DOCUMENT RESUME

ED 397 533 EA 027 929

AUTHOR	Khattri, Nidhi; And Others
TITLE	Assessment of Student Performance. Volume II: Case Studies. Studies of Education Reform.
INSTITUTION	Pelavin Research Inst., Washington, DC.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	1 Nov 95
CONTRACT	RR91172004
NOTE	439p.; For Volumes I-III of this particular study, see EA 027 928-930. For all 12 final reports (36 volumes) in this series of studies, see EA 027 926-961.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC18 Plus Postage.
DESCRIPTORS	Academic Achievement; Accountability; Case Studies; Data Analysis; Data Collection; *Educational Assessment; *Evaluation Criteria; Instructional Improvement; Longitudinal Studies; *Organizational Objectives; *Performance; Secondary Education; Student Evaluation; Teaching Methods
IDENTIFIERS	*Studies of Education Reform (OERI)

ABSTRACT

The emphasis on using performance assessments to support systematic state-, district-, or school-wide purposes is a recent development in the assessment-reform movement. This document presents data from a 3-year case study that was designed to elucidate the status of assessment reform in United States education systems and offer recommendations for policy and future research. Researchers made visits to 16 schools during spring 1994; they revisited 7 of the sites during spring 1995. Data were derived from documents, telephone interviews, observations, and onsite interviews. Each case-study summary contains four sections: a school profile and introduction, a description of performance assessment, the context of implementation, and the use and impact of performance assessment in the school. Appendices contain a sample mathematics task; a list of essential skills for reading, writing, and mathematics in grades 9-12; and mathematics rubrics. (LMI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 397 533

STUDIES OF EDUCATION REFORM: ASSESSMENT OF STUDENT PERFORMANCE

FINAL REPORT

VOLUME II: CASE STUDIES

Project Director: Michael B. Kane

Authored By:

Nidhi Khattri
Alison L. Reeve
Kerry Traylor
Rebecca J. Adamson
Douglas A. Levin

November 1, 1995

Submitted To:

Office of Educational Research and Improvement
U.S. Department of Education

Submitted By:

Pelavin Research Institute
1000 Thomas Jefferson Street, Suite 400
Washington, DC 20007

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

EA027929

BEST COPY AVAILABLE

STUDIES OF EDUCATION REFORM: ASSESSMENT OF STUDENT PERFORMANCE

FINAL REPORT

VOLUME II: CASE STUDIES

Project Director: Michael B. Kane

Authored By:

Nidhi Khattri
Alison L. Reeve
Kerry Traylor
Rebecca J. Adamson
Douglas A. Levin

November 1, 1995

Submitted To:

Office of Educational Research and Improvement
U.S. Department of Education

Submitted By:

Pelavin Research Institute
1000 Thomas Jefferson Street, Suite 400
Washington, DC 20007

PREFACE

Studies of Education Reform: Assessment of Student Performance is the result of a research project conducted by Pelavin Research Institute (PRI), an affiliate of the American Institutes for Research (AIR), under a contract with the Office of Educational Research and Improvement (OERI) of the U. S. Department of Education (Contract Number RR91172004). In 1991, OERI issued a request for proposals entitled "Studies of Education Reform." Twelve studies were incorporated under this general heading, each reflecting some aspect of the reform movement that had placed education at the forefront of the national agenda in the late 1980s. PRI was awarded a three-year contract to study assessment reform, which we interpreted to mean the contribution of performance-based, non-multiple choice assessments to education reform.

Contractors for all 12 reform studies were required to hold a national conference within the initial year of their study and to commission papers on important aspects of the reform topic. PRI, in collaboration with the OERI study of curriculum reform (conducted by Ronald Anderson of the University of Colorado), held a national conference on performance assessment and curriculum reform as a pre-session to the Annual Student Assessment Conference, organized by the Education Commission of the States, in Boulder, Colorado, in June of 1992. The assessment component of the pre-session conference included discussions of the content of nine commissioned papers that are to appear in the book, *Implementing Performance Assessments: Promises, Problems, and Challenges* (Kane & Mitchell, in press).[1] The papers, the conference attendees' insights, and OERI's research questions helped us refine our study's intellectual and methodological framework.

The larger and more significant context for this study was the increasing commitment across the nation to performance assessment as a reform strategy. For example, California spearheaded the reform movement with statewide open-ended mathematics assessments in the late 1980s, and Vermont followed suit with its first, statewide portfolio assessments. In the late 1980s and early 1990s, other states, districts, and schools also began developing and implementing performance-based assessments.

In the subsequent two years, we visited 16 schools across the country (of which we revisited seven) that were participating in the development or implementation of performance assessments as a result of national, state, district, or local assessment reform initiatives. In addition, we presented papers based upon our initial study findings at the American Educational Research Association conference in April, 1995, and the Council of Chief State School Officers' conference on Large-Scale Assessment in June, 1994 and 1995.

This volume, the second in a three-volume series, presents case studies of the 16 schools we visited across the country. These schools were involved in developing and implementing performance assessments as a result of national, state, district, or local assessment reform initiatives. The results of the project are contained in Volume I: Findings and Conclusions, and the specific research objectives and study design are presented in Volume III: Technical Appendix — Research Design and Methodology.

---

[1] All royalties resulting from the sales of this book will be contributed to the Leigh Burstein Memorial Fund, administered by the University of California at Los Angeles Foundation.

ACKNOWLEDGMENTS

The support and participation of several individuals and institutions made this study possible. We are grateful to the state and district education agency personnel and to the principals, teachers, students, parents, and school board members who devoted large amounts of their precious time to answer our questions and to guide us in understanding the issues related to assessment reform. We regret that our promise of anonymity leaves them without the public recognition they deserve.

We also are grateful to David Sweet, the OERI monitor of *Studies of Education Reform: Assessment of Student Performance*, whose incisive feedback guided our work at every step, and to Ruth Mitchell, who served as Project Manager during the study's initial stages and as Senior Advisor during the later years. Her steady interest in and support of the project was invaluable.

This study also benefitted from the participation and general support of the members of our Advisory Committee, including: Dr. George Elford, Educational Testing Service (retired); Dr. Lauren Resnick, Learning Research and Development Center, University of Pittsburgh; Dr. Ramsey Selden, Council of Chief State School Officers (now of the Education Statistical Services Institute); Dr. Karen Sheingold, Educational Testing Service; Dr. Loretta Shepard, Department of Education, University of Colorado; and Dr. Grant Wiggins, Center on Learning, Assessment, and School Structure. In 1991-92, the Advisory Committee members also included: Dr. Robert Mislevy, Educational Testing Service, and Mr. Sol Pelavin, American Institutes for Research.

We also wish to acknowledge the capable support and contributions of the following individuals: Ray Varisco, Gwen Pegram, Kimberly Cook, Michael Garet, Kimberly Gordon, Rebecca Shulman, and Shelley Kirkpatrick, all at Pelavin Research Institute; and Amy Stempel and Stepahnie Soper, both at the Council for Basic Education.

TABLE OF CONTENTS

PREFACE i

ACKNOWLEDGMENTS iii

TABLE OF CONTENTS v

INTRODUCTION 1

Double-Visit Case Studies

SITE	SITE VISIT DATES
Coalition of Essential Schools: Cooper Middle School	April 7-8, 1994 March 23-24, 1995
Primary Learning Record: Park Elementary School	May 5-6, 1994 April 3-4, 1995
Language Arts and Math Portfolios: Niños Bonitos Elementary School	April 18-19, 1994 February 27-28, 1995
Applications Assessments: Westgate Middle School	June 6-7, 1994 February 16-17, 1995
Vermont's Portfolios: Maple Leaf Middle School	May 10-11, 1994 March 16-17, 1995
Oregon Assessment Reform: Crandall High School	March 11-12, 1994 April 17-18, 1995
Kentucky Assessment Reform: Breckenridge Middle School	May 3-4, 1994 March 9-10, 1995

Single-Visit Case Studies

SITE	SITE VISIT DATES
Rite of Passage Experience: Thoreau High School	April 12-13, 1994
The College Board's Pacesetter Mathematics Program: Sommerville High School	May 10, 1995
The New Standards Project: Ann Chester Elementary School	May 25-26, 1994

Single-Visit Case Studies, continued

SITE	**SITE VISIT DATES**
The New Standards Project: Noakes Elementary School	February 21-22, 1995
Harrison School District 2's Performance-Based Curriculum: McGary Elementary School	April 11-12, 1994
South Brunswick's Sixth Grade Research Performance Assessment: Windermere Elementary School	June 2-3, 1994
New York Regents Portfolios: Hudson High School	May 12-13, 1994
Maryland Assessment Reform: Walters Middle School	May 9-10, 1994
The Arizona Student Assessment Program: Manzanita High School	April 27-29, 1994

INTRODUCTION:
AN OVERVIEW OF STUDY OBJECTIVES AND DESIGN

Objectives

The major objectives of the three-year longitudinal study ***Studies of Education Reform: Assessment of Student Performance*** are as follows:

Objective 1: *Document and analyze key characteristics of performance assessments;*

Objective 2: *Document and analyze facilitators and barriers in assessment reform; and*

Objective 3: *Document and assess impacts of performance assessments on teaching and learning.*

Our ultimate purpose in this study was to elucidate the status of assessment reform in U.S. education systems and to offer recommendations for policy and future research.

Research Design[1]: Case Studies

Our research design employed a qualitative, case-study approach to collecting data about performance assessments and their impacts at the school level.

We designed a modified time-series approach for gathering data, which enabled us to obtain both cross-sectional and longitudinal data. Cross-sectional data allowed us to make comparative remarks about assessments and school contexts. The longitudinal data allowed us to document the effects of and changes in performance assessments over time within sites.

We selected 16 sites (the definition of "site" for this study encompasses both a performance assessment and a single school at which it is being used) which a team of two researchers visited a single time during a two-day site visit. We then selected a subset of 7 sites, which the team returned to for a second visit (therefore, longitudinal data were collected for only 7 of the 16 sites).

We conducted the first set of site-visits in the Spring of 1994, and the second set of site-visits in the Spring of 1995. (Two of the single time site-visits were conducted in the Spring of 1995). Exhibit 1 shows our site-visit design.

---

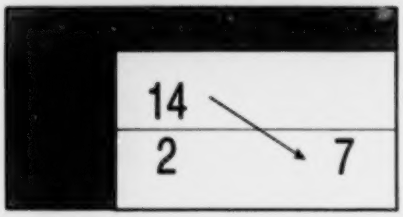
[1] A full discussion of study objectives and design is continued in Volume III: Technical Appendix — Research Design and Methodology.

Sample Selection Criteria

As described above, our research design called for two waves of data collection: first round of visits to all sites included in the sample, followed by a second round of visits to a subset of those sites. Below, we describe the criteria we applied to select sites for inclusion in rounds one and two.

EXHIBIT 1

Site Visit Design



Selection Criteria: Round One Sites

The overarching objective of our site selection process was to identify, insofar as possible, a set of school sites that exhibited the range of experiences American schools are having with the development and implementation of performance assessments. For the purposes of our study, we defined a case study site as a single school where a performance assessment was being implemented. To select the sites, we delineated two sets of criteria – those pertaining to performance assessments and those pertaining to schools.

Performance assessments are marked by a number of variable characteristics, and we attempted to obtain variation in our sample within each characteristic. Selection criteria pertaining to performance assessment characteristics included:

- **Type of Assessment**. Performance assessments come in a variety of forms, including portfolios, on-demand assessments, demonstrations and presentations, and extended projects. We wanted our sample to include assessments that reflected this range so that we might discern variation in effects of assessment type on teaching and learning at the local level.

- **Locus of Development**. The movement toward the use of performance-based assessments is taking place at all levels of educational authority. States, districts, and schools alike are developing and implementing performance assessments. Furthermore, some national-level efforts, such as the New Standards Project and the Coalition of Essential Schools, are also influencing the turn toward performance assessments. Because the purposes, design, and impact of assessments developed at different levels of authority could potentially vary significantly, we wanted our sample to reflect this diversity in locus of development.

- **Status of Implementation.** The entire performance assessment movement is still relatively young. However, performance assessments do vary with respect to their stage in development and implementation. Therefore, we wanted our sample to include assessments with varying status of implementation: developmental and pilot, full-scale implementation, and maintenance.

- **Content Area.** Performance assessments can be used in all subject areas, but the assessments can look different for different subject areas. We wanted our sample to include assessments that focused on a range of subject areas including language arts, mathematics, science, and social science.

We attempted also to obtain variation across two school characteristics:

- **School Level.** We wanted the sample to include elementary schools, middle schools, and high schools, since performance assessments might affect teachers and students differently at the various levels of schooling.

- **Geographical Diversity.** Because American children are educated in schools in 50 states and the District of Columbia, and because these schools are located in urban, suburban, and rural areas, we wanted the sample to include school sites located in various regions of the country and in communities of varying urbanicity.

Selection Criteria: Round Two Sites

We chose a subset of seven sites for a second round of data collection. We based selection of the seven round two sites upon one or more of the following criteria:

- It was anticipated that changes in the performance assessment design or implementation would take place between 1993-94 and 1994-95.

- Our understanding of the effects of assessment reform at the site was less than clear based on one round of data collection and was likely to improve with a second visit.

Sample Description

Sixteen performance assessments at 16 school sites were selected to comprise the study sample. The 16 sites are identified in Exhibit 2.

EXHIBIT 2

16 Round One Sites*

SITES

State-level assessments:

- Arizona Assessment Reform, Manzanita High School
- Kentucky Assessment Reform, Breckenridge Middle School
- Maryland Assessment Reform, Walters Middle School
- New York Regents Portfolios, Hudson High School
- Oregon Assessment Reform, Crandall High School
- Vermont Portfolio Assessments, Maple Leaf Middle School

District-level assessments:

- Harrison School District 2's Performance-Based Curriculum, McGary Elementary School (CO)
- South Brunswick Unified School District's 6th Grade Research Performance Assessment, Windermere Elementary School (NJ)
- Prince William County Public Schools' Applications Assessments, Westgate Middle School (VA)

School-level assessments:

- Language Arts and Math Portfolios, Niños Bonitos Elementary School (CA)
- Primary Learning Record, Park Elementary School (NY)
- Rite of Passage Experience, Thoreau High School (WI)

National-level assessment projects:

- New Standards Project, Ann Chester Elementary School (TX)
- New Standards Project, Noakes Elementary School (IA)
- Coalition of Essential Schools, Cooper Middle School (NM)
- College Board's Pacesetter Mathematics Program, Sommerville High School (MD)

* All schools have been assigned pseudonyms.

Data Collection Activities and Instruments

Because we were interested in obtaining information about the performance assessment, the educational context within which it was developed and implemented, and the assessment's effects at the local level, we collected documentary, phone interview, and site-visit data.

Documentary Data

Prior to and during each site visit we collected background documentary data about the subject assessment. The available data varied across assessments. Types of data collected include:

- Descriptions of performance assessments,
- Samples of performance assessments,
- Policy documents about the assessments,
- Policy documents about related education reform efforts,
- Evaluation and research reports regarding the assessments, and
- Newspaper reports about the assessments.

These data were collected from state and local education officials, school staff members, and representatives of external groups involved in assessment reform (e.g., the New Standards Project). These data were collected throughout the life of the project.

We also collected documentary data about the school sites we visited. These data included reports describing each school's demographic composition, staff description, financial resources, and other relevant documents.

Phone Interview Data

Prior to each site visit, we also conducted initial telephone interviews with cognizant individuals in state and local education offices, the school site, and external assessment reform organizations. We used an interview protocol tailored to the role of the interviewee and to the performance assessment system under investigation.

Site Visit Data

In the Spring of 1994 we visited the first 14 schools in our sample. In the Spring of 1995 we revisited seven schools and added two new ones to our sample. In total, we conducted 23 site-visits.

Each site visit lasted one-and-a-half to two days and was conducted by a team of two researchers. The researchers interviewed a number of individuals, observed classrooms, and, whenever possible, observed professional development sessions devoted to the development and use of performance assessments, administration of performance assessments, other activities related to the implementation of performance assessments.

We used semi-structured interview protocols during our site visits. The protocols for both waves of data-collection were quite similar in structure, but wave-two protocols contained more probing questions about the use and effects of performance assessments on teaching and learning.

Case Study Data Presentation

Each case study write-up is divided into the following four sections:

- **Section One: School Profile and Introduction.** This section briefly profiles the organizational characteristics of the school, such as demographic data about its students and the community it serves. The section also presents information on the number and roles of the individuals interviewed and the types of observational data collected.
- **Section Two: Description of Performance Assessment.** This section presents a brief history of the development and implementation of the performance assessment. It also presents the key characteristics of the assessment, including purposes, content areas assessed, scoring procedures, and technical characteristics.
- **Section Three: Context of Implementation.** This section includes a summary of the policies and procedures followed in implementing the performance assessment, the resources and help available to the education agency or school personnel for developing and implementing the assessment, and the coordination (or lack thereof) between performance assessments and other tests, reforms, and organizational changes.
- **Section Four: The Use and Impact of Performance Assessment.** This section describes the uses of the performance assessments by teachers and students at the sample school. In addition, it documents the school community's evaluation of the usefulness and quality of the assessments and the impact of the assessments on the teaching and learning processes at the school.

Case Study Data Interpretation

A research design such as the one we used has strengths, but it also necessarily imposes certain limitations on the interpretations that can be drawn from the data. We briefly outline general limitations of our case study approach to data collection.[2]

First, although we attempted to obtain a representative sample of performance assessments, we are not certain that the assessments initiated at the district and school levels are, in fact, representative of all district- and school-initiated performance assessments.

Second, our information regarding the facilitators and barriers in assessment reform, especially at the national-, and state-levels may be less comprehensive than for those at the district- and school-levels. This limitation stems from the local-level emphasis of our study. We collected information regarding national- and state-level assessment reform from documents and general, as opposed to detailed and probing, interviews. In addition, we did not conduct in-person interviews with state officials and researchers involved in national-level efforts as we did with district- and school-level personnel.

---

[2] Specific limitations are discussed in Volume I: Findings and Conclusions and Volume III: Technical Appendix — Research Design and Methodology.

Third, our findings regarding the impact of national-, state-, and district-initiated performance assessments are valid only for the schools included in this study; the results obtained for a particular school cannot be generalized to other schools involved with the same performance assessments.

Finally, interviewees' opinions regarding impact of and problems with performance assessments signal the existence of those impacts and problems, but the absence of such opinions does not necessarily suggest the absence of impact of or problems with performance assessments.

Coalition of Essential Schools:
Cooper Middle School
April 7-8, 1994
March 23-24, 1995

COALITION OF ESSENTIAL SCHOOLS: COOPER MIDDLE SCHOOL

Introduction

The focus of this case study is the development and implementation of performance assessments at a member-school (Cooper Middle School) of the Coalition of Essential Schools (the Coalition). Cooper Middle School, like the other Coalition schools, develops its own performance assessments, based on the principles and philosophy of education expounded by the Coalition as a whole.

Cooper Middle School is located in Santa Fe, NM. The scenic beauty of the Sangre de Cristo mountains, the distinctive adobe architecture, and ethnic and cultural diversity fuse to lend a special flavor to this southwestern city. Tourists flock to revel in its attractive ambience, and artists have long captured its compelling colors in various mediums. Behind its attractive facade, however, the community is troubled.

Community schools, according to study participants, are witnessing tensions between whites ("Anglos") and Hispanics; high schools are experiencing a student drop-out rate of around 11.4 percent; and the area is suffering from student gang violence. Furthermore, the community's lack of respect for the teaching profession is a significant hurdle continually confronted by the public school system. It is against the backdrop of such troubles that Cooper Middle School is attempting reforms to better meet the needs of its students and to prepare them to become lifelong learners.

Reflecting the community at large, 54 percent of Cooper Middle School's approximately 630 students are white, 45 percent are Hispanic, and less than 1 percent are Native American. The school spans grades seven and eight, and its graduates advance to the two local public high schools (and some attend the private schools in the area). About a third of its students qualified for Chapter I funds in 1993-94, but the figure dropped down to 25 percent in 1994-95, due to redistricting and to the opening of a new middle school. The state-wide figure for Chapter I eligible students stood at 26 percent in 1994-95.

Participants

Several individuals were interviewed for this case study, as illustrated in Exhibit I.

EXHIBIT I

Study Participants

1993-94	1994-95
• Coalition of Essential Schools Researcher • Principal • Twenty-two Teachers – 10 seventh grade teachers – 12 eighth grade teachers • Two eighth grade students • One local NEA Representative • One parent • One School Board Member	• Principal, Cooper Middle School • Three teachers – one seventh grade – one eighth grade – one elective • Eight students – 4 seventh grade students – 4 eighth grade students • One local NEA Representative • Two parents • One School Board Member

Observations

In addition to conducting these interviews, in 1994 the study staff observed a student-led mock court session and in-class instruction, and in 1995 they observed three in-class periods.

This report first outlines the Coalition's educational and assessment principles and then reviews their manifestation and impact at Cooper Middle School.

Coalition of Essential Schools

The Coalition of Essential Schools was established in 1984, at Brown University, as a school-university partnership to help "redesign" schools. As of December 1, 1993, Coalition members included 150 schools that are actively involved in school reform. The work of member schools is guided by a set of nine *Common Principles*, intended to help initiate a *conversation* about important issues in education and to provide a broad framework for reform activities (see Appendix A).

Assessment Reform

The sixth of the nine common principles pertains to assessment, and states that students should be awarded a diploma only upon successfully demonstrating — through an *exhibition* — that they have acquired the skills and knowledge central to the school's program. *As the diploma is awarded when earned, the school's program proceeds with no strict age grading and with no system of `credits earned' by `time spent' in class. The emphasis is on the students' demonstration that they can do important things* (*The Common Principles of the Coalition of Essential Schools*, Appendix A).

The Coalition suggests that a process dubbed *planning backwards* is essential to all aspects of reform, including assessment reform. *Planning backwards* entails first developing a vision for the school and then taking the steps to realize that vision (ergo, "planning backwards"). Joe McDonald, a senior researcher at the Coalition, has enumerated the steps in planning backwards as follows:

- **Defining a vision**. To define a vision, the school sets aside any curriculum and testing requirements, and generates visual images — that is, *exhibitions*, of what it expects of its graduates. Exhibitions ". . . originate as articulations of a school's vision — movies for a mind's eye, portraying graduates using *their* minds well."[1] These visions are ". . . concrete images of real kids" that stimulate teaching and at the same time function as an assessment tool.

- **Building a platform**. In building a platform, the school compares what its students are able to do with its own ideal vision of a graduate. In comparing the actual performances to the ideal, the school uses its assessment system as the pivotal tool of reform. Based upon real versus ideal comparisons, the school begins to change its structures and processes to come closer, in reality, to the ideal vision.

- **Rewiring**. This step entails overhauling the structures and process in a major way — rewiring the system, so to speak, based upon the questions and information emerging from the second step (building a platform). Rewiring includes steps such as . . . *the invention of new junctures at which, for example, teachers come together to exchange graded papers and discuss grading standards, or at which parents and teachers and members of the larger community come together to share perspectives on achievement, or at which classes combine across grades and levels to discuss the same texts or problems or to get involved in some activity.* The method centers on the act of taking the steps to invent a culture of *inquiry* and *collaboration*.

- **Tuning**. Continual *tuning* of the rewiring process is viewed as critical to keeping alive the school's commitment to self-examination and improvement. An important aspect of tuning entails extending the information network to include the community at large.

Thus, as the school approaches its vision, it develops a more extensive information network that incorporates all stakeholders. Through this information network, the school includes the community in its reform efforts and benefits from the information it receives from the world outside its physical boundaries.

Research and Dissemination

Through its newsletter, *Horace*, and through other publications, information from the Coalition's extensive research and development activities is disseminated widely to member schools and to other interested audiences. For example, the Autumn 1993 list of publications includes titles such as *Steps in Planning Backwards* by Joseph McDonald and *An American Teacher's view of British Assessment Practices*, by Marshall Cohen. In addition, the Coalition has produced an electronic publication entitled *The Exhibition Collection* that contains samples of exhibitions developed by member schools, including samples of student work and assessment standards. The Coalition also holds regional and national conferences and *round-table discussions* centered around a variety of reform concerns — such as exhibitions of student work and pedagogical strategies.

---

[1] Joseph McDonald, (February, 1992). Steps in planning backwards. Coalition of Essential Schools, Brown University. pg.1.

In addition to such dissemination efforts, the Coalition also works with *RE:Learning*, a partnership that includes the Education Commission of the States (ECS) and all member schools, districts, and states. *RE:Learning* provides financial and professional help for reform efforts to its members.

School Context

Cooper Middle School's assessment reform is embedded within, and is an integral component of, its larger systemic reform program. Cooper joined the Coalition and launched its comprehensive reform program in 1989, at which time the school realized that it could no longer operate as a ". . . mini high school and hope to meet the needs of early adolescents."[2] Because of its reform efforts, the school was selected in March 1991 as the New Mexico Middle School of the Year.

Through much of this time, the school essentially struggled on its own, without much professional or systemic support and involvement from the district, but this state-of-affairs changed somewhat during the 1994-95 school year. In 1994-95, the district became interested in the portfolio system, and Cooper is now in the position of helping one of the local High Schools develop portfolios for its ninth grade students. As a result of this cooperation, Cooper students' portfolios will be accepted by the area high schools, when Cooper students advance to the high school level.

Two other major district-level changes affected Cooper as well. The district acquired a new Superintendent, who, according to several study participants, will seek greater accountability from the schools. Cooper teachers, thus, already are contemplating the alterations they might have to make in the future to their home-grown curriculum and assessment strategies. In addition, Cooper's boundaries changed, requiring semi-adjustment on the parts of teachers and the Principal.

School Mission and Reforms

Cooper's mission statement is as follows:
Students who graduate from Cooper Middle School will have the critical thinking and social skills to develop into positive, responsible, and productive citizens who can envision how they can contribute to the society in which they find themselves. They will be empowered with the confidence to be independent, resourceful,

---

[2] *Cooper Middle School*, November, 1993, pg. 1.

committed and successful lifelong learners who will see their individuality and that of others in a positive light.[3]

As a result of the reform process launched in 1989, the school reorganized itself into *family groups* composed of a core group of teachers and students. Through the 1993-94 academic year, each of the two eighth grade *families* were composed of six teachers and a heterogeneous group of students, and each of the two seventh grade families were composed of five teachers and a heterogeneous group of students. (All special education students were *mainstreamed*, excepting 12 students classified in the "D" category.) The elective teachers comprised a distinct, STAR, family, which is responsible for teaching subjects such as music, arts, home economics, foreign languages, and computers to both 7th and 8th graders.

For the 1994-95 academic year, however, the family structure was altered to form three teams of three teachers in the seventh grade, and two teams of three teachers and one team of six teachers in the eighth grade. These structural changes were decided upon during a school planning session in the Summer of 1994. Cooper staff came to believe that the extant family structure had not functioned well during the previous year, because teachers had formed families based on personal compatibilities, rather than on academic requirements. In one family, for example, none of the teachers had felt comfortable teaching mathematics and science, and, therefore, the family could not provide its students rigorous training in those two subject areas. In addition, teachers also believed that smaller family sizes would be more conducive to planning, implementing, and evaluating the family-based curriculum and assessment systems.

The structural changes discussed above did not affect the functional features of the family. Each family plans its own curriculum, schedule, and student counseling services. Reportedly, coordination of school programs is done through a "shared decision-making process" consisting of the principal and a representative from each family (or, regarding certain matters, the entire family).

Following the Coalition philosophy, each family presents a large part of its curriculum materials focused on *essential questions that lead students to final assessments centered on their abilities to use critical thinking skills in order to solve real-world problems.*[4] The school has spelled out the critical thinking skills requirement into several *exit competencies*, which students are required to demonstrate upon graduation. These competencies, developed by a task force of teachers and parents, incorporate state and district competency requirements. The focus on exit competencies is provided through thematic units integrating skills, knowledge domains, instructional strategies, and assessment systems. (Two samples of thematic units are presented in Appendix B.) All thematic units and assessments are designed by teachers themselves and are organized around *essential questions.*

The seventh grade thematic units are generally organized around the essential question *Who am I?* — and the eighth grade units are formulated around *How do I affect society?* Each family plans 4 to 10 thematic units and associated assessment techniques per year. Assessments, thus, are an integral and an on-going feature of curriculum and instruction.

---

[3] Ibid, pg. 2.

[4] Ibid, pg. 2.

In 1994-95, in addition to changing the family structure, there also was a change in the extent to which teachers used thematic units for content area coverage. As will be discussed under the Impact on Curriculum and Instruction section, teachers placed less emphasis on thematic units as the preferred pedagogical strategy, and this change was the most significant for mathematics.

The development and implementation of assessments are described in the sections below.

Performance Assessments

The school fashioned performance assessments to: (a) better capture students' learning with regard to the *essential questions* and the *integrated curriculum*, (b) to inform the student about his or her progress throughout the school's program, and (c) to inform parents about their child's interests and progress in school.

The school requires students to keep portfolios of their work, to undergo the *Rite of Passage Experience* (ROPE), and to answer an *Open-ended Interdisciplinary Question* at the culmination of eighth grade. These three comprehensive assessments build upon the numerous projects and assignments — *performance tasks* — students complete over the course of their two years at Cooper.

It is important to note, however, that despite the use of a family-based integrated curriculum and assessment system, the school continues to use a traditional school report card that is sectioned into traditional subject areas. Students are assigned letter grades based not only upon their project work, essays, and reports, but also upon multiple-choice tests, and none of the comprehensive assessments — a complete portfolio, ROPE, and the *Open-ended Interdisciplinary Question* — is a district requirement for graduation.

The development of project-based performance assessments and their characteristics, as well as the comprehensive assessments themselves, are described below.

Development of Performance Assessments

Reportedly, assessment development was, and continues to be, inspired by the Coalition's method of *planning backward*. Thus, outcomes are conceptualized first, and curriculum, instruction, and assessment tasks or assignments are fashioned around the resulting outcomes. That is, teachers first delineate the skills and knowledge they want their students to be able to demonstrate at the end of a unit and then design thematic units around those skills and knowledge outcomes. Elements of *planning backwards* were evident in teachers descriptions of the steps taken to plan for the 1994-95 academic year.

During the Summer of 1994, and in subsequent teacher forums, teachers discussed the successes and failures of the previous year in terms of student performance. For example, they evaluated ROPE performances and noted that students rarely mentioned mathematics in their presentations. As a result, in 1994-95, Cooper teachers began taking a more direct, traditional approach to the teaching of mathematics. However, none of the teachers interviewed for this study detailed the processes of *planning backwards* in creating the assessment formats and instruments, and assessments and curriculum were generally seen as being tightly interwoven.

Each family plans and coordinates its own thematic units, which vary in format and completion time. For example, the eighth grade *Plaid Flamingos Family* teachers work together on *The Dawn of the Nuclear Age Unit*, which lasts for about four and a half weeks (see Appendix B). The essential questions and skills for the unit are spelled out, as are the assignments and projects on which the student is to be assessed.

Assignments and Projects

Assignments and projects consist of essays, student productions, debates, or other events that are based on a student's completion of a research project. For example, during the Spring of 1994, eighth grade students had to select and write an essay on a topic dealing with *The Intersection of Medicine and Science* (see Appendix C). Students were asked to research the topic, cover enough information in their reports to convey the topic to their readers, pay attention to the "quality and mechanics" of their writing, and use visual aids, such as graphs and charts, in their reports. An example of another assignment, a student production, is described in Exhibit II, below.

EXHIBIT II

Court of Law

An eighth grade family held a mock court trial as the culminating activity of a thematic unit entitled ***Freedom and Responsibility***.

The topic of the mock trial was, *Should the United States of America ban the sale, manufacture, and use of cigarettes?* One Judge, a prosecutor, a defense attorney, witnesses, a jury, a transcriber, and a court marshall, all played by eighth grade students, constituted the "court." About 42 students sat in the audience and one student videotaped the proceedings. One at a time, the lawyers put forth well-informed arguments, produced expert testimony and witnesses, cross-examined one another's witnesses, and delivered closing statements. After a brief deliberation, the jury rendered its verdict. The court proceeding lasted about fifty minutes.

At the conclusion of the mock trial, all students were required to produce a persuasive piece of writing based upon the court proceedings, to be incorporated into individual student portfolios.

Scores on assignments and projects like the above are determined through scoring rubrics. Sometimes one teacher scores one entire assignment or project, while at other times several teachers score various parts of an assignment, so that grades for the traditional subject areas can be determined.

Scoring Rubrics

The design of each scoring rubric is determined by the nature of the assignments and projects and by the expertise of the teacher who designs it. Each scoring rubric outlines the dimensions of the assignment to be graded and the score distribution or the maximum score value for each dimension. For example, for an essay, *What is a Hero?*, the scoring rubric established three dimensions: *Writing Skills*; *Content*; and *Format* (see Appendix D).

Each major dimension was further divided into sub-dimensions: the *Writing Skills* dimension consisted of spelling, punctuation, and sentences; the *Content* dimension consisted of the exposition of common qualities of heroes and the explanation of those common qualities; and, finally, the *Formant* dimension consisted of an introduction and a conclusion. The *Writing Skills* and *Content* sub-dimensions were to be rated on a 5-point scale, and the *Format* sub-dimension were to be rated on a 3-point scale. The teacher-marked scoring rubrics are returned to the students for feedback about their performance and as an instructional tool for further work.

Comprehensive Assessments

In addition to the Family designed projects and assessment practices, each student is expected to compile some of her or his scored projects and assignments into portfolios, to participate in the *Rite of Passage Experience* (ROPE), and to answer an *Open-ended Interdisciplinary Question*. These standard school-wide assessments build on one another to provide a cumulative record of the student's academic work at Cooper, and are used to inform students and their families of student progress with regard to critical thinking and social skills. The teacher who is responsible for the student's portfolio and ROPE communicates the information to the student and to his or her family. However, as previously mentioned, none of these comprehensive assessments is required for graduation, and none is reflected on the student's report card.

Portfolios

All seventh and eighth grade students are required to maintain a portfolio that each edits during the last term of eighth grade. This portfolio must include:

- Projects or assignments that highlight the student's work in the areas of Language Arts, Math, Science, and History and Social Studies;
- One piece of creative or expressive work that reveals something about the student;
- A best piece of work from an arts class; and
- A two-to-three page essay on the student's academic, physical, and social growth during the year.

Each student edits her or his portfolio during the last term of her or his eighth grade year. Next, each student's *family of teachers* examines the portfolio and assesses her or his critical thinking skills, interests, and areas of strengths and weaknesses, and provides the student with written feedback. In the seventh grade, the portfolio is not used to provide such written feedback. Although the portfolio does not receive an overall score at the end of the year, it is primarily used for overall self-evaluation during the last nine weeks of school.

For the first time, at the end of the 1994-95 academic year, eighth grade students also will winnow their portfolios and send their exemplary portfolio pieces to their future high school.

Rite of Passage Experience (ROPE)

All graduating eighth graders must undergo a *Rite of Passage Experience* (ROPE). Students design and present to a panel of teachers, parents, and community members an audiovisual master portfolio of their experiences and growth while at Cooper. The panel acts as a ". . . sounding board for the students' self-reflections and self-appraisals."[5] ROPE typically lasts for 15 minutes, and a record of the ROPE conference is sent to the student and to his or her family. (The ROPE report format, again, depends upon what the family wants to do.)

Interdisciplinary Open-ended Question

Students write an answer to an *interdisciplinary open-ended question* at the time they enter seventh grade, and then again during the spring semester of eighth grade. A comparison of the student's seventh grade essay with his or her eighth grade essay enables the student to appraise his or her growth over the two-year period; the student's eighth grade answer is incorporated into his or her final portfolio. Each family develops its own rubric to evaluate the *interdisciplinary open-ended question.*

Resource and Professional Development Support

Cooper participated in a fair number of professional development activities in the past few years. Prior to the school joining the Coalition, Cooper Middle School's Principal went to visit the Coalition's headquarters at Brown University to review some programs and to meet with researchers and program directors. The Coalition, in turn, sent researchers to help teachers with *planning backward* for curriculum, instruction, and assessments. In addition, for training and professional development, teachers visited schools in San Diego, Kentucky, and New York City, which are experimenting with developing and implementing performance assessments. The visiting teachers specialized in specific assessment and instructional strategies, such as portfolios and cooperative learning.

By 1993-1994 the focus of professional development activities had shifted almost entirely to the process of whole-staff decision-making. Seven teachers visited other schools specifically to study decision-making processes. In addition, the school added five minutes of instructional time per period (at its own discretion). It also provided five days in-service for the year, three of which were used to "*action plan*" changes for overall school reform. Few teachers, thus, remembered receiving in-house professional development geared towards the development and use of performance assessment strategies. During this year, the Coalition provided information on assessments and related issues primarily through publications.

During both years teachers emphasized their need for in-class training and support for the express purpose of designing and using assessments and thematic units. They repeatedly said that they wanted mentors to be working beside them to provide them with hands-on, authentic training and experience. Teachers noted that they often have to help themselves, and, although they are quite resourceful, the process can be quite burdensome. For example, the special education teacher felt he did not fully understand how to use portfolios and that he required greater support in the design and utilization of what he called "true" portfolios. He did not remember any professional development over the past three years that explicitly centered on performance assessments. In fact, the teacher observed, one of the greatest impacts of the reform efforts has been the need for "intense professional development," geared toward classroom practices. The Principal was aware of the situation and hopes to make extensive use of in-house expertise in the coming academic year. He plans to encourage teachers to spend time with other families in the school.

---

[5] *Cooper Middle School*, 1994, pg. 5.

In the meantime, to facilitate thematic teaching and the use of performance assessments, teachers generally schedule 45 minutes per day for team-planning and another 45 minutes for personal use. According to the Principal, teachers generally set aside their personal time on Tuesdays and Thursdays for parent-student-teacher conferences, on Mondays and Wednesdays for curriculum planning, and on Fridays for "other things." He said that the teachers generally spend about two to three hours per week for assessing how their students are performing on thematic units.

Cooper's professional development activities discussed above were supported by the New Mexico RE:Learning Partnership and the Panasonic Foundation, who provided $5,000 and $2,000 respectively per year. Begun in 1988, the RE:Learning partnership comprises the Coalition, the Education Commission of the States (ECS), and all member states, district, and schools. Its aim is to assist systemic reform efforts at any level of the education system.

According to the Principal, the RE:Learning grant in 1993-94 was used for professional development and for reimbursing substitute teachers at $50 per day. In 1994-95, it was used for professional sessions devoted to evaluating Cooper's old competencies and for developing new ones. The Panasonic funds in 1993-94 were used for keeping school statistics, such as student attendance rates, and in 1994-95, the Foundation helped pay for professional development and planning sessions focused on evaluating Cooper competencies.

For obtaining such financial and in-kind support, Cooper Middle School submitted proposals explaining what they would do with the funding. They must identified extant needs in the school and how the money and the resources would help fulfill those needs.

Interaction with Other Reforms and Tests

Other Reforms

As mentioned previously, the assessment reform was and is part and parcel of an overhaul of the school's organizational structure, curriculum, and instructional strategies. Supposedly, assessment techniques and results are used to drive changes in instruction and curriculum through the concept of *planning backwards*, but most interviewees suggested that during the initial stages of reform, the conceptualization and implementation of these changes were close to simultaneous.

A major outcome of the overall reform effort was the institutionalization of a goal-setting process for new students called *Two Weeks into Your New Experience* (TWINE). During the TWINE conference, the individual student, his or her parents, and the teacher talk about the school's program and each person's expectations for the academic year TWINE, thus, functions as a goal-setting mechanism for ROPE and also as a method of inviting parental participation in Cooper students' education.

In 1994-95, although *planning backwards* as a technique of systemic change was not identified as such by most study participants, "exhibitions" of student performance and data from other assessments were used to evaluate and change the *family* structure and pedagogical strategies.

Other Tests

In addition to its own comprehensive and on-going performance assessments, the school administers, each spring, the Iowa Test of Basic Skills (ITBS) to eighth grade students, including all Chapter I and Special Education students. The ITBS is required by the district. According to the teachers and to the Principal, district-wide scores on ITBS have not increased in the past year, but Cooper's scores have been higher than those of other middle schools. In the 1993-1994 testing period, Cooper students scored at the 64th national percentile in reading, 49th in math, 50th in social studies, and eighth in science. Figures similar to those were obtained in 1992-1993.

In 1993-94, teachers said that they did not alter their instructional strategies to prepare students for the ITBS, and, ostensibly, there was no coordination between the school's curricular, instructional, and assessment strategies (to match the multiple-choice format) and the administration of ITBS.

In 1994-95, however, as will be discussed in greater detail under Curriculum and Instruction, teachers did change their overall pedagogy strategies for mathematics and science, because Cooper students' ITBS results were not satisfactory. Teachers, however, disliked the ITBS, as they felt that it does not have an acceptable structure for assessing the kinds of skills and competencies they are trying to instill in their students. Students, too, disliked the ITBS and find it to be "useless." In fact, one student said that she did not perform well on the ITBS, because it contained material she had not yet been exposed to in her mathematics classes.

Impact of Assessment and Other Reforms

No formal studies have evaluated the impact of the myriad reforms at Cooper. However, the school community is taking note in detail of the complexity of the effects upon itself of such changes. Such effects are discussed below.

Impact on Teachers

Although Cooper teachers support the reform activities and outcomes, the "transition has been difficult," said the Principal. Teachers shoulder the demanding tasks of fashioning their own thematic units and assessments to judge student achievement. For them, this has raised, rather than resolved, several issues. (Many teachers, especially in math, continued to use traditional, multiple-choice tests in addition to performance assessments for the purposes of evaluating student performance and assigning traditional letter grades to their students.)

Quality and Standards. For many teachers, among the other main points to be debated are assessment *quality* and *standards*. In 1993-94, some of the teachers were worried that the quality of their assessments, i.e., reliability and validity, might be questionable; by 1994-95, this worry extended to the appropriateness of their teaching methods as well. In 1994-95, teachers also observed that it had become increasingly difficult for many of them to be generalists and to rotate the teaching of several subject areas.

One teacher verbalized the difficulty as feeling ". . . very frustrated." He said that teacher needed to explore other models of thematic teaching and special education inclusion practices. Through the thematic units they designed and the rubrics they generated and refined every year, they maintained certain expectations of and objectives for their students, but the larger educational objectives remained obscure. (In 1993-94, for example, some teachers were not fully aware of the school's defined competencies, and, by 1994-95, new competencies were being developed.) The NEA teacher representative stressed that outcomes needed to be clearly defined and sequenced, if the assessments were to be useful.

Report Card. Another major dilemma teachers continued to face is that the report card they must use is organized by traditional discipline areas, and certain skills and competencies must be covered within those subject areas. Assignment of subject area letter grades based on interdisciplinary thematic units is, therefore, quite challenging. As one teacher said, teachers are doing "two dances — [the] innovative and [the] required dance," which was a frustrating situation. (As mentioned above, part of a student's grade in the traditional subject areas also derived from these assessments.)

In order to assess and report the attainment of Cooper competencies in a uniform format, the Cooper staff worked on designing a rubric-style report card during 1994-95. The Principal hoped to use this report card (in addition to the traditional report card) at the end of the 1994-95 academic year.

Professional Support. Teachers also talked about needing information on how to score entire projects and wanting on-going technical assistance, more money, and more time to devise assessment strategies and thematic units. They could not really discuss assessment issues without addressing instructional concerns, and, hence, stressed that the drawbacks of project-oriented instruction and assessment were that there was very little time to "refine" units and assessments. Teachers have about 45 minutes per day for planning, and that is too little to discuss the benefits and drawbacks of the new instructional and assessment techniques.

Impact on Curriculum and Instruction

Cooper teachers regularly used a variety of instructional techniques, ranging from engaging students in Socratic discussions, cooperative learning, debates, and lectures accompanied by audiovisuals. The principal maintained that the teachers use more varied instructional strategies now than they did before the reforms were mobilized. Positive educational changes, however, both teachers and principal agreed, cannot be driven solely by revamping assessments; issues in curriculum and instruction must also be considered.

In 1993-94, some teachers were concerned that thematic units might interfere with sufficient coverage of traditional discipline areas. According to the NEA representative, student choice in determining thematic assignment topics precluded comprehensive coverage of content for every student. By 1994-95, this concern was particularly salient for comprehensive coverage of mathematics, science, and history.

In the Summer of 1994, using ROPE as an informational tool, teachers identified what they perceived to be weaknesses in their mathematics curriculum. They were startled to discover that students hardly ever mentioned mathematics in their culminating performances, ROPE, either as something they had learned or as something with which they had experienced difficulty. Teachers also were disappointed in their students' ITBS mathematics scores. They came to the conclusion that teaching mathematics mostly through applied problems and thematic units were not adequate methods of instilling in their students a good understanding of the discipline. These teachers felt that their methods had not been successful either because the methods had been developmentally inappropriate for their students, or because they did not possess sufficient mastery in using those methods. In either case, teachers had experienced "constant frustration" of teaching mathematics as an applied science, said one teacher.

Cooper teachers, thus, decided to employ a more direct skills-building, "old fashioned" approach to the teaching of mathematics. During 1994-95, therefore, teachers placed less emphasis on teaching mathematics through thematic units and a little more stress on "pure" mathematics. Every student, thus, was assigned to a fifty minute mathematics class. One eight grade family also decided to group its students by ability levels for its mathematics classes. In addition, to further understand the issues involved in mathematics curriculum, a school committee was established to study the NCTM standards to further refine the school's mathematics program.

Cooper teachers also refined other disciplinary areas as well as thematic units. During 1994-95, in fact, teachers identified *all* disciplines by their names, and they also tried to emphasize traditional teaching methods. According to one teacher, part of the pressure came from the parents, who asked, "What class is this?" (It also was evident from the parent-teachers conferences that parents felt that their children were not being taught sufficient amounts of mathematics and science and that these areas had to be strengthened.) Thus, for example, American History was presented in a more chronological fashion, and science was infused even more into thematic units and was taught as a separate subject.

Teachers of elective subjects said they had always used performance assessments, but that they now had written standards and benchmarks (but no scoring rubrics) to judge student performance. The electives teachers also said that assessment reform was concentrated primarily in the "core" classes, as there was not much need for such reform in their classes. The technology teacher said she supports reforms in the "core" classes by requiring students to play electronic games that promote inductive and deductive reasoning and problem-solving skills.

Impact on Students

In 1993-94, according to the two students interviewed, thematic units were enjoyable *and* confusing. The two mentioned that students were uncertain about which teacher taught what subject, a factor that caused them some distress. For example, one student mentioned that he would have liked more help in mathematics, because he was not proficient in the subject, but that he did not know who to approach for help. (He aspired to study civil engineering in college, but needed a rigorous math background do be able to do so.)

Another student said that she did not like the thematic units because they did not "teach subjects." (Because of her elementary school training, she was accustomed to being exposed to extensively defined subject areas.) In addition, both students were worried about gaining content area knowledge, because the report card, the basis for graduation, was organized by subject area.

A related disgruntlement these students harbored was that they did not understand how subjects were interrelated within thematic units, or what, exactly, they were expected to know at the end of those units. Students articulated the desire for ". . . more organization of themes — everything is mixed up in themes; we don't know where we are." One student claimed that he was "lost" when they covered the essential question, "*What does it mean to be human?*"

Students also complained that they found thematic units burdensome, as they were asked to complete research-based assignments, even though they frequently were not given a bibliography or resources, and despite the fact that the library did not contain much research material. These students also did not evince strong feelings about portfolios, because it was ". . . not a big thing," according to one. That student said that it was only "partly kids' choice" as to what completed assignments were to be included in the portfolios. One student, in fact, kept using the word "miscommunication" about the process of education at Cooper. Paradoxically, both students did not mind creative assignments, and one said that he especially liked debating.

By 1994-95, many of these difficulties had been allayed, primarily as a result of clearer identification of traditional subject areas and due to the changes in the family structure. Most students reported that their classes were "pretty separated," and they knew when they were studying mathematics, science, English, and other subjects. Their reactions to thematic units also were more favorable. They said that they enjoyed doing "hands-on" assignments, such as monitoring the stock market for mathematics and making "dragsters" for science projects. They also said that they liked keeping portfolios, because they can revisit the work they had done. These students also enjoyed group work when they did not have to finish other students' work, which did happen from time to time.

Students' reactions to the grading system showed a preference for the traditional letter-based grading system, perhaps because that is the basis for Cooper's report card. In 1993-94, although students appeared not to be overly concerned with the grading system, because, as one said, it was "fair, more or less," they mentioned wanting more feedback and a better understanding of the grading criteria. They also said the grades seemed to them based on accumulating "things," rather than on the "quality" of what they produce. In 1994-95, students were clearer about the use of scoring rubrics for setting performance standards and for scoring purposes. Nonetheless, one seventh grade girl said that she liked the traditional letter-grade system, because an "*'A'* shows me that I'm doing really good."

In 1993-94, both the principal and the teachers reported that as a result of the classroom teaching and learning strategies employed at Cooper, their students were developing powerful critical thinking skills, good writing skills, the motivation to achieve high standards, and the ability to work together groups. The eighth grade teachers observed that their students were much more challenged through the project work than through text-book based work, and were focusing on "processing information." Interestingly, teachers from one *family* asserted that the one direct result of assessment was that scoring rubrics helped students understand the grading criteria, so that they no longer disputed their grades.

At the end of the 1993-94 academic year, however, teachers also noted that students were not proficient enough in mathematics and in science. In addition, because of the team structure, teacher rotation of subject areas, and the use of thematic units, students often were confused about which teacher taught what subject and about the connections between thematic units and subject areas. As a result of this realization, the team structure and pedagogical strategies underwent significant changes. These changes are discussed in the sections below.

Special Education Students. The special education teacher asserted that performance assessments was "wonderful" for special education students, but such assessments, especially exhibitions and portfolios, tended to reward more "overt and gregarious" children. He said that children who are emotionally and behaviorally disturbed did especially well on performance assessments because they relied on verbal skills to express themselves. The special education teacher also believed, however, that other types of assessments that place less emphasis on "show" must be devised to capture the performance of students who are shy.

Impact on Parents and Community

The school is regarded as a leader in assessment reform and enjoys a good reputation within the community. In 1993-94, according to a school board member, the board was "very pleased and supportive of it [the school]." In fact, the Board was trying to encourage other schools to reform their instructional and assessment methods. However, the board member also was concerned about the lack of specific standards and the communication of those standards to students. (He, too, was not aware of the Cooper competencies.)

The board member's personal concern about his daughter who attended the school involved projects and homework. He did not "see a real structure to her projects," and he felt that students were assigned "little homework." Hence, in his opinion, the right objectives were in place, but not the right processes and outcomes.

In 1994-95, the school board member also had a similarly high opinion of Cooper, but he had no children enrolled in the school. His information, thus, was more pertinent to the district as a whole. He was concerned about the district's ability to attract quality staff, as the district's financial incentives were so low. He also expressed the need for establishing a rigorous district-wide accountability system, as he was distressed about the fact that the district scored on the lower end of ITBS. Although he expressed great enthusiasm for using portfolios as means of giving feedback to students, he did not want them to be converted into a "machine-like process."

In both years, all parent participants were supportive of Cooper's philosophy of education and teachers use of project-based work, but two parents also expressed concerns over children's exposure to content areas. "Method for method sake is meaningless. Transferral of content information is very important," said one parent.

In 1993-94, the parent interviewee was an active member of the school community. At Cooper, she coached a group of students for an academic competition entitled "Odyssey of the Mind." Although she was appreciative of the hands-on work and creativity demands of project-oriented work at Cooper, she was not certain that her son liked the work very much. She mentioned, too, that ROPE is a great idea, but she is not so sure that it means as much as it is supposed to mean to the students. She, herself, was concerned about what she perceived to be a weak scoring system, and she wanted a more rigorous one in place.

In 1994-95, while one parent, the President of the Parent Teacher Association, spoke very highly of the school, the other parent felt that Cooper standards must be raised and learning outcomes must be defined. (He had not yet seen Cooper's new competencies.) Both parent were supportive of the plans they said the new district Superintendent has to institute greater accountability measures for the school system.

Future Plans

In the near future, Cooper staff plans to address a number of outstanding and important issues. It will (a) devise a better inclusion model for special education students, (b) require an educational plan for every student, and (c) define a better mode of transition from middle to high school.

Cooper teachers currently feel that Cooper students experience more difficulties adjusting to high school classes than do students from other middle schools, because they become accustomed to project work and performance assessments. The changes in their approach to teaching subject areas, they hope, will help the future cohorts in adjusting to the academic demands at the high school level.

The school also will finalize new Cooper competencies and better assessment and measurement of student outcomes. The new competencies will be established within the state's curricular frameworks entitled, *Standards of Excellence.* In the meantime, the school has developed a Cooper Competency form that is intended to function as a narrative report card. Students will eventually be rated on Cooper Competencies at the end of seventh and eighth grades. The Principal also hopes eventually to use the form for advancing or detaining students. In the meantime, the district's new Superintendent may also require changes in assessments and curricular frameworks.

The school administration also would like to move from *time-bound learning* to *mastery-based education* in order for students to determine their own pace in mastering curriculum units. However, standards for competencies and performances have not as yet been fully articulated, and faculty members are experiencing conflicts over how to meet all the currently established standards. They feel that they cannot provide variable amounts of time to students to attain mastery of the various units.

All of the above plans will be subject to the scrutiny of the new district superintendent. The common wisdom is that the new superintendent will standardize the curriculum and seek more accountability. It is the hope of the Cooper community that the district will spell out the required competencies, but that it does not require all schools to use specific textbooks and lesson plans. Cooper Principal also hopes that Cooper will be used as a model school for reform activities in the district.

Conclusions

Cooper is steeped in the philosophy of school reform and is earnestly attempting to change the central features of education: teaching and learning. The school has taken great pains to reorganize itself into smaller student and faculty groups (*families*) and is grappling with issues in the areas of curriculum, instruction, and assessment. The faculty uses the information from student performance to judge the rigor of its pedagogical approach and philosophy. Problems, however, remain.

There appears to be little central coordination in the development and implementation of thematic units and the accompanying assessment system. In addition, less than needed professional development in designing curriculum, assessment, and instructional strategies means that teachers are left to their own devices, which adds work to already overloaded schedules. Hence, several features of the assessment system are uncoordinated and without much quality control. Teachers tend to use "existence proofs" (e.g., the presence of certain ideas, characteristics, and so on) as grading criteria; and standards for student outcomes have not been fully defined.

Coupling interdisciplinary instruction and assessment with a traditional content area report card also is a cause for concern. Both teachers and students worry about content coverage and about how to *infer* performance in a subject area from an interdisciplinary assignment.

Fundamentally, the issue of quality in assessment reform centers around the resources (basically information and time) provided to teachers for continuing the work of reform (without causing them to become confused about the "two dances" they are expected to perform).

Cooper staff is aware of these issues and concerns, and, in 1994-95, began to address them with changes to its family structure and pedagogy. The questions that led instructors to change their approach to teaching, however, remain unanswered.

APPENDIX A

The Common Principles of the Coalition of Essential Schools

THE COMMON PRINCIPLES OF THE COALITION OF ESSENTIAL SCHOOLS

1 • The school should focus on helping adolescents **learn to use their minds well.** Schools should not attempt to be "comprehensive" if such a claim is made at the expense of the school's central intellectual purpose.

2 • The school's goals should be simple: that each student **master a limited number of essential skills and areas of knowledge.** While these skills and areas will, to varying degrees, reflect the traditional academic disciplines, the program's design should be shaped by the intellectual and imaginative powers and competencies that students need, rather than necessarily by "subjects" as conventionally defined. The aphorism "Less Is More" should dominate: curricular decisions should be guided by the aim of thorough student mastery and achievement rather than by an effort merely to cover content.

3 • The school's **goals should apply to all students,** while the means to these goals will vary as those students themselves vary. School practice should be tailor-made to meet the needs of every group or class of adolescents.

4 • **Teaching and learning should be personalized** to the maximum feasible extent. Efforts should be directed toward a goal that no teacher have direct responsibility for more than 80 students. To capitalize on this personalization, decisions about the details of the course of study, the use of students' and teachers' time and the choice of teaching materials and specific pedagogies must be unreservedly placed in the hands of the principal and staff.

5 • **The governing practical metaphor of the school should be student-as-worker,** rather than the more familiar metaphor of teacher-as-deliverer-of-instructional-services. Accordingly, a prominent pedagogy will be coaching, to provoke students to learn how to learn and thus to teach themselves.

6 • Students entering secondary school studies are those who can show competence in language and elementary mathematics. Students of traditional high school age but not yet at appropriate levels of competence to enter secondary school studies will be provided intensive remedial work to assist them quickly to meet these standards. **The diploma should be awarded upon a successful final demonstration of mastery** for graduation — an "Exhibition." This Exhibition by the student of his or her grasp of the central skills and knowledge of the school's program may be jointly administered by the faculty and by higher authorities. As the diploma is awarded when earned, the school's program proceeds with no strict age grading and with no system of "credits earned" by "time spent" in class. The emphasis is on the students' demonstration that they can do important things.

7 • **The tone of the school should** explicitly and self-consciously **stress** values of **unanxious expectation** ("I won't threaten you but I expect much of you"), of **trust** (until abused) and of **decency** (the values of fairness, generosity and tolerance). Incentives appropriate to the school's particular students and teachers should be emphasized, and parents should be treated as essential collaborators.

8 • **The principal and teachers should perceive themselves as generalists first** (teachers and scholars in general education) and specialists second (experts in but one particular discipline). Staff should expect multiple obligations (teacher-counselor-manager) and a sense of commitment to the entire school.

9 • Ultimate administrative and budget targets should include, in addition to **total student loads per teacher of 80 or fewer pupils, substantial time for collective planning by teachers, competitive salaries for staff and an ultimate per pupil cost not to exceed that at traditional schools by more than 10 percent.** To accomplish this, administrative plans may have to show the phased reduction or elimination of some services now provided students in many traditional comprehensive secondary schools.

Any list of such brevity and specificity begs for elaboration, and it is this elaboration that must first engage the energies of each Essential school. The process of designing programs and putting them into place will take several years, and the inevitable adjustments then required will consume some years after that. And due to the need to adapt each design to its own constituency of students, teachers, parents and neighborhoods and to create a strong sense of ownership of it by those who are involved, this redesign must be largely done at the level of the individual school — even as that school adheres to the principles and standards common among the Coalition's member schools.

APPENDIX B

Sample Thematic Unit

MOONGAZERS

OVERVIEW

STUDENTS WILL DEMONSTRATE THEIR UNDERSTANDING OF CONF
IN THE DEVELOPMENT OF NEW MEXICO DURING THE 19TH CENTUR

ESSENTIAL QUESTION

HOW DID CONFLICT SHAPE NEW MEXICO?
WHY DID PEOPLE COME TO NEW MEXICO?
WHAT CHANGES RESULTED FROM THE CONFLICT IN NEW MEXICO?
WHAT IMPACT DID THE OLD SANTA FE TRAIL HAVE?

THEME

19TH CENTURY NEW MEXICO

ESSENTIAL SKILLS

The student will be introduced to the following skills:

- writing a paper and delivering a narrative. IIBlc
- interpreting and relaying verbal and nonverbal directions/messagesIIA6
- asking questions to elicit specific information/opinions IIA7
- use a variety of sources to distinguish relevant and irrlevan information IAla
- recognize social, historical, political, cultural and environme relationships IA5b
- reconize the impact that the events of the past have on the future. IOA5d
- develop sense of chronology of events IAlc
- acknowledge sources and references through use of bibliographies. IB
- compare and contrast various kinds of information. IIA4
- write a paper and deliver an oral presentation which is cle consise, and interesting in each of the following modes informative, narrative, analytical. IIBla, c, d
- Write using standard grammar, spelling , punctuation, by editing and proof reading work IIB2
- Show respect for self, adults, peers, and property IIIC
- Generate ideas and formulate hypothesis to predict outco IA2
- Perform all four operations with real numbers and formulas to solve math problesm IA4a
- Identify bias, propaganda and faulty logic and distinguish between fact and opinion. IIA5

BEST COPY AVAILABLE

MOONGAZERS

BODY OF KNOWLEDGE/PEDAGOGY

JOURNALS
MONOLOGS
POETRY
ISOLATION OF NEW MEXICO
SANTA FE TRAIL
METHODS OF COMMUNICATION
SOLVING PERCENTAGE PROBLEMS
USING PERCENTS TO DISCRIBE AND PREDICT POSSIBLE OUTCOMES
HOW PERCENTS PERSUADE US
NOTE-TAKING SKILLS
WRITING BIBLIOGRAPHIES
ESSAY FORMAT
COMPARATIVE WRITING
VOCABULARY DEVELOPMENT
OUTLING FORMAT AND SKILLS
RESEARCH
ORAL PRESENTATION
MAPPING
TIMELINES
CUENTOS
DICHOS

BENCHMARK OUTCOMES

ENGLISH	Essay, Outline, Bibliography, Note Cards Create Individual book
SOCIAL STUDIES	Note Cards, Rubric, Picture, Map, Timeline
Math	Estimating distances, To-Scale Drawings
COMMUNICATIONS	Monologs, Journals, Poetry, Multi-Med Presentations

ASSESSMENT

The family will pick one:

1. **PRODUCTION**: Students will generate their material for th own production for "The Santa Fe Trail"

2. **INDIVIDUAL BOOK** Each student will write his/her c cuento of their family arriving in NM

3. **OLD SANTA FE TRAIL JOURNAL** Each student will make replica of an Old Santa Fe Trail Journal and write his/her diary of travel along the Old Santa Fe Trail

BEST COPY AVAILABLE

THE DAWN OF THE NUCLEAR AGE

Level:	Eighth grade
Overview:	Students will study the technology, history and issues surrounding the development of nuclear energy and participate in a debate.
Theme:	Nuclear Energy (a 4 1/2 week thematic unit)

Essential Questions:

1. Was the U.S. justified in using the Atomic Bomb on Japan?
2. Should the government continue to support the design and manufacture of nuclear weapons?
3. How safe is nuclear energy?
4. How was the U.S. government justified in using uninformed citizens for radiation testing during the Cold War?

Bodies of Knowledge:

Science	Nuclear reactions
	Genetics and mutations
	Growth and Human Development
Social Studies	History and Impact of nuclear discovery
	Effects on foreign policy
	World War II; the Cold War
Mathematics	Probability
	Equations
Language Arts	Readings- fiction, nonfiction
	Research
	Historical Fiction, writing and reading
	Public speaking/Debate

Vocabulary

Genetics, radiation, fission, fusion, heredity, radioactivity, atom, atomic, nuclear, the Cold War, McCarthyism, propaganda, explosion, implosion, totalitarianism, dictator, racism, appeasement, facism, Marxism, Nazism, isolationism

D.O.N.A DEBATES

STUDENT NAME ____________________________________

DEBATE RESOLUTION:____

________(5 pts.) The student turns in a complete script of her/his 3 minute debate before performing the debate.

________(20 pts.) The Debate Topic is clearly understood. Relevant terms are defined (If first positive or first negative, there is a brief history given.) .

________(20 pts.) The positive's or negative's main idea is made clearly and the debater demonstrates convincingly that his/her position is valid and necessary.

________(5 pts.) Each debater uses at least one quote which supports the argument. Source of quote must be noted.

________(10 pts.) All factual information is given a credible source.

________(10 pts.) Charts or graphs effectively display useful and meaningful information.

________(10 pts.) Style and delivery of presentation.

________(10 pts.) Effective use of props (Bonus pts.)

REBUTTAL

________(10 pts.) The rebuttal is clear, well organized and well thought out.

________(10 pts.) The rebuttal effectively points out weak points in the opponents argument and/or successfully strengthens weak points in the debater's argument.

________(100pts.) TOTAL

APPENDIX C

The Intersection of Medicine and Science Project

Modern Developments in Medicine and Science

Due: Friday, April 15–100 points

THE ASSIGNMENT:

Select a topic from some point at which science and medicine intersect (their "nexus"). This topic should be of interest to you personally. Research the subject and write a report about what you have learned about this subject.

In reporting on your topic, include as much specific information as possible about how it affects the human body and tell which systems and organs are affected. Include also information about prognosis for the future and possible breakthroughs in treatment. If you are reporting about such controversial medical issues as the abortion pill or AIDS explain why this subject is controversial and what the sides of the controversy are.

Included with your report must be some sort of **visual aid** that helps "fill out" your report for those of us who need visual stimulus in order to learn. For example, if you are reporting on AIDS you could display through pictures or drawings how the virus harms the white blood cells, or you could chart the sequence of how Cancer spreads through tissues.

Here are a few suggestions if you can't think of anything:

1. There are always articles in TIME, NEWSWEEK, and PEOPLE magazines and in the newspapers about breakthroughs in medical research. Find at least 3 articles that deal with research about one disease or problem, read them, and write a brief summary of the articles, relating the information to what you are learning about the human body and its systems. Include the articles with the rest of your project.
2. There is a constant flow of information about the AIDS virus. Find out more about how the virus works in the body. Find photos or drawings of HIV. Display in pictures how the virus harms the white blood cells. Chart a sequence of what happens to the body's immunity to other diseases. Give some information about the latest in HIV research and list ways in which AIDS is and is not spread.
3. There are many different kinds of acne medications and skin lotions for sale. Most of the information that you get about these products is from newspaper or magazine ads, and radio and TV commercials. Before you buy you should evaluate a product's ad claims. Do an evaluation by following these steps: (1) Find and cut out a newspaper of magazine ad for an acne medication or skin cream or lotion. (2) Write the name of the product. (3) Read the ad carefully, then make a list of questions you would like to have answered before you decide whether or not to believe what the ad claims. (4) What reasons might you have for questioning the claims in the ad? (5) What in the ad made you want to buy the product? (6) What in the ad seemed not to be completely truthful? (7) Would you buy and use the product in the ad? Explain your answer. (8) Design your own ad for a skin care product that effectively "sells" the product.
4. On your own research briefly, and write a report on some other news from the nexus of science and medicine. E.g. report on Bo Jackson's hip replacement or other sports medicine stories; cancer research and treatment like the use of Shark Cartilage; the necessity of vitamin and mineral supplements in the diet; the dangers of skin cancer as a result of exposure to the sun; the hazards of drug / alcohol abuse to the body; etc...

You will be graded on:
1. Evidence of research (be sure to cite your source -- minimum of two)
2. Quality and mechanics of your report --spelling, grammar, punctuation, neatness of presentation, etc...
3. Thoroughness of the report--do you cover enough of the information to make the report of use and interest to others who may be curious about the subject
4. Visual aid(s)--graphs, charts, pictures, drawings etc...

Primary Learning Record: Park Elementary School

May 5-6, 1994
April 3-4, 1995

PRIMARY LEARNING RECORD:
PARK ELEMENTARY SCHOOL

Introduction

This case study explores the integration of a new method of assessment with a well-developed and well-articulated educational program and philosophy with which it is compatible. Educators at New York City's Park Elementary have long provided a child-centered educational program, and for the past few years, several of them have employed the British-developed *Primary Language Record*, and, more recently, the *Primary Learning Record*, to help support that program.

Park Elementary is a magnet school in New York City's District Four. Founded 20 years ago, the school's guiding philosophy is to provide a child-centered education. The school has waivers from some district mandates, including one that allows teachers to send home narrative reports in lieu of standard report cards. During the 1993-94 school year, Park Elementary served about 250 students, pre-kindergarten through grade six. These students were 18 percent white, 36 percent African-American, 42 percent Hispanic, and 4 percent Asian. (The demographics of the student body were little changed for the 1994-95 school year.) The Park Elementary staff includes primarily experienced teachers; about two-thirds of the staff of 14 have been at the school for 15 years or more, and several teachers have been at the school since its founding. All regular classroom teachers teach mixed-grade classes (e.g., K-1, 3-4), and students typically have the same teacher two years in a row.

The *Primary Language Record* (PLR) provides teachers with a structured method of tracking young children's development of language skills and planning individualized instruction to meet students' language and literacy needs. The PLR was first introduced in New York City in the early 1990s. The PLR has since spawned the Primary Learning Record (PLeR), which uses the same approach to monitor and record children's intellectual development in all subject areas. Implementation of first the PLR and, beginning in the 1994-95 school year, the PLeR in about 62 New York city schools is supported by three organizations — the Center for Collaborative Education, the Center for Educational Options, and the Elementary Teachers Network — which together comprise the New York City Assessment Network.

This case study is drawn from two visits to Park Elementary, the first in May 1994 and the second in April 1995. Between the two visits, teachers had switched from using the PLR to the PLeR. However, their reactions to and comments about the two closely related assessments are quite similar. Thus, throughout this case study, the combined term, "PLR/PLeR" will be used to designate facts or opinions applicable to both versions of the assessment.

Participants

Several individuals associated with Park Elementary and the Center for Collaborative Education (the organization assisting Park Elementary teachers in the implementation of the PLR/PLeR) were interviewed in Spring 1994 and Spring 1995; the roles of these individuals are identified in Exhibit 1.

Observations

In the Spring of 1994, lessons were observed in two classrooms (preK-K and 5-6) and samples of completed PLR forms were collected. In Spring 1995, lessons were observed in two classrooms (preK-K and K-1).

EXHIBIT I

Study Participants

1993-94	1994-95
• Teacher Consultant at the Center for Collaborative Education • Teacher Director • Four Teachers (grades preK-K, 3-4, 5-6, and special education resource); three of the four teachers use the PLR • One parent	• Teacher Consultant at the Center for Collaborative Education • Three teachers (grades preK-K, K-1, and special education resource); all three teachers use the PLeR • One parent

School Context and the Primary Language and Learning Records

Park Elementary is housed in an educational complex in East Harlem, which also houses Park Elementary's associated high school (grades seven through twelve) and a neighborhood junior high school. However, the wing that houses Park Elementary feels very much like an elementary school. The hallways are lined with children's work — paintings and drawings, sculpture and crafts, stories and discoveries. Children seem to feel comfortable interacting with all of the school's teachers, calling all staff members by their first names.

Families must apply to have their children admitted to Park Elementary; in order to apply, the child and family members must first spend several hours visiting the school. Admissions are then made by lottery, with the exception of children of staff members and siblings of other Park Elementary students, who are automatically admitted.

The Center for Collaborative Education and the New York City Assessment Network

Park Elementary is one of 11 schools participating in the Center for Collaborative Education's (CCE), Elementary School Assessment Project (ESAP). CCE, a network of small alternative schools in New York City, initiated ESAP in the early 1990s. The mission of ESAP is "to develop the use of qualitative assessment techniques for instructing and evaluating children's learning in the New York City public school system. . . . using highly detailed observations and descriptions to capture the richness of children's learning".[1]

---

[1] Wallace, V. (1993, May). *Elementary School Assessment Project: Review of Year One, 1992-93.* New York: Center for Collaborative Education, p. 1.

In 1991, the CCE joined with two other New York City organizations, the Center for Educational Options and the Elementary Teachers Network, to establish the New York City Assessment Network (NYAN). The three organizations formed NYAN because of their shared commitment to introducing to their teacher colleagues an assessment technique called the *Primary Language Record* (PLR), a tool designed to help teachers acquire and use those techniques espoused in the ESAP mission. In the Fall of 1994, NYAN and its participating schools shifted from the PLR to the PLeR. All together, the three NYAN organizations are supporting approximately 300 to 400 teachers in about 62 schools in New York City as they implement the PLR/PLeR.

The Primary Language Record and the Primary Learning Record

As designed, the PLR is intended to accomplish two major purposes: (1) to help the child's classroom teacher understand the child's learning style — focusing specifically upon language development — in order to tailor an educational program appropriate for the child; and (2) to help the teacher communicate his or her understanding of the child to the child's next teacher. The PLeR serves the same purposes and incorporates the same components as the PLR, but it is intended for use by teachers to monitor children's development in all subject areas, not just language.

Teachers at Park Elementary use the PLR/PLeR to help them better understand their students' learning styles, literacy progress, and other intellectual development and to plan instructional activities to meet the needs of individual students. Teachers do not use the PLR/PLeR to assign grades or to rank children on their progress. Thus, as the PLR/PLeR is implemented at Park Elementary, the first purpose identified above is more important than the second.

Development of the Primary Language Record and the Primary Learning Record

Both the PLR and the PLeR were developed by British educators associated with the Inner London Education Authority (ILEA) and led by the Centre for Language in Primary Education. Starting in 1985, these educators identified a need in their schools and classrooms of gaining a better sense of their students' learning strategies and their emerging literacy; they identified this need in part because of the growing number of children in ILEA schools for whom English was not a first language. Piloted in 50 schools during the 1986-87 school year, the PLR was made available in September 1988 to all British schools interested in using it.

As developed, the record aims to satisfy both record keeping purposes and instructional purposes, including:

- Informing teachers who do not yet know the child of his or her progress;
- Informing the principal about the child's progress;
- Providing parents with information about the child's progress; and
- Supporting and informing teaching in the classroom and providing teachers with a framework for teaching language and literacy.

The last objective is the one of primary importance to NYAN, CCE, and teachers at Park Elementary as they use the PLR and the PLeR (also developed by ILEA educators).

Introduction of the Primary Language Record and Primary Learning Record to New York City and Park Elementary

In the early 1990s, several of the British educators involved in the development of the PLR spoke in New York City about their experiences with it. Representatives from all three NYAN member organizations attended the presentation and became interested in introducing the PLR to the teachers they worked with. One Park Elementary teacher attended the presentation as well, and she began experimenting with the PLR during the 1991-92 school year. Other teachers at Park Elementary did not begin using the PLR until 1992-93, the year the Center for Collaborative Education's Elementary School Assessment Project began. Teachers at Park Elementary are not required to use the PLR or to participate in ESAP, but about half of the school's teachers do so voluntarily.

In the Fall of 1994, CCE and all Park Elementary teachers involved in ESAP switched from using the PLR to the PLeR. The switch to the PLeR was motivated by feeling among teachers that it was more natural to observe and record children's learning in all areas, not just in language and literacy development. All New York City teachers collaborating with the CCE switched to the PLeR at this time.

The Role of NYAN Teacher Consultants

Each of the three organizations involved in NYAN — the Center for Collaborative Education, the Center for Educational Options, and the Elementary Teachers Network — has at least one *teacher consultant* on its staff. This individual is responsible for helping teachers at its member schools become comfortable with the PLR and use it successfully. These teacher consultants foster implementation of the PLR in several ways, by:

- Convening group discussions with teachers to speak of specific children's learning styles and how instructional and curricular approaches can be tailored for children;
- Conducting in-depth observations of individual children to share insights about the children's learning with their teachers; and
- Helping teachers find ways to adapt the PLR so that it fits naturally into their teaching style and becomes an asset.

CCE's teacher consultant works with six teachers at Park Elementary — about half of the Park Elementary staff. (As a former Park Elementary kindergarten teacher, the teacher consultant is not a stranger to Park teachers. Additionally, CCE occupies as its office one large room in the same educational complex in which Park Elementary is located; thus CCE staff are readily accessible to Park Elementary teachers.) During 1992-93, the CCE teacher consultant also worked with 33 teachers at nine other schools; she typically visited schools about once a week. By 1994-95, despite the change from the PLR to the PLeR, teachers had grown more accustomed to and adept with the PLR, and, consequently, the teacher consultant was able to reduce the frequency of her visits to schools participating in ESAP. During the 1994-95 school year, she visited schools about once a month. All teachers — both at Park Elementary and other CCE member schools — participated voluntarily with CCE and the PLR/PLeR.

Characteristics of the PLR and PLeR

The PLR was developed primarily to help teachers develop the language and literacy skills of children in the primary grades. However, the PLR can be used with older children as well, and Park teachers also used it in other subject areas (thus demonstrating the viability of and paving the way for the *Primary Learning Record*).

The Primary Language Record

The developers of the PLR recognized that teachers are not the only individuals in children's lives who have insights into their language development. Thus, the PLR process is designed specifically to draw upon multiple perspectives of a child's development of language skills. To be as comprehensive as possible in its record, the PLR process:

- Involves parents in conferences with the teacher to comment on the child's language development at home;
- Records the child's own evaluation of his or her progress; and
- Includes information about a child's language development in his or her first language, if that language is not English.

The PLR consists of two types of record forms, the Primary Language Record form and the Observations and Samples form:

- The ***Primary Language Record Form*** has sections for the teacher to record in the fall and spring of the school year. In the fall, the teacher records (a) information about the languages spoken and read by the child and the child's medical history, (b) notes from a conference with the child's parents about the child's language development, and (c) notes from a conference with the child about his or her language skills; and in the spring, the teacher records cumulative observations about (d) the child's listening and talking practices and development, (e) the child's reading development, (f) the child's writing development, (g) notes on a follow-up conference with the child's parents, including the parents' response to the teacher's comments on parts d, e, and f, (h) notes from a follow-up conference with the child, and (i) information for the child's next classroom teacher.
- The ***Observations and Samples Form***, on which the teacher notes: (a) the child's talking, listening, reading, and writing behaviors and the contexts in which those behaviors occur; (b) the titles of texts read by the child and notes about the child's reading progress and reaction to the particular text; and (c) information about the child's written work, including the context in which the writing occurred (e.g., alone or in a group, with consultation from the teacher or without, type of writing, and the child's and teacher's response to the writing). Notes taken on this form are used by the teacher to plan future lessons and to aid in the end of the year

summary of the child's development on the Primary Language Record form. Use of the forms is flexible and does not involve applying a scoring scheme to student performance; rather, teachers adapt the forms as they see fit and use them according to their own schedule.

Copies of the Primary Language Record Form and the Observations and Samples Form appear in Appendix A.

The Primary Language Record Handbook for Teachers is a 65-page guide to help teachers use the record forms and learn to interpret the records they keep.[2]

The Primary Learning Record

The PLeR uses the same types of forms as the PLR except that they are designed to help teachers structure and assess their observations of children's learning in all subject areas. As the PLeR is used by teachers participating in ESAP, the forms themselves are largely irrelevant (they use PLR and PLeR forms interchangeably; indeed, when asked if they could provide copies of the PLeR forms, the teacher consultant and several Park Elementary teachers produced PLR forms instead): the switch from the PLR to the PLeR was motivated by teachers' decisions to expand the scope of their recorded observations, not so much by the forms themselves.

Examples of the PLR and PLeR in Use

The teachers at Park Elementary who have used the PLR and who are using the PLeR describe the stages of inquiry they go through as they focus on a child, using the PLR to guide and record their observations. These stages of inquiry include:

- How did the activity the child is involved in originate (e.g., Did the teacher assign or suggest an activity? Did another child initiate the activity? Did the child choose the activity himself or herself?)
- What did the child do during the activity?
- What questions did the child raise during the activity?
- What did the child do when he or she got stuck?
- Where did the child go with the activity?
- What should the teacher do next to build on the child's learning experience with the activity?

In the Spring of 1994, two teachers at Park Elementary and the CCE teacher consultant independently explained the same (or virtually the same) set of questions they ask themselves as they observe the children they teach.

---

[2] Published by Heinemann Educational Books Inc. Note that, to date, the Handbook and forms remain in their original British formats. Consequently, some of the terminology used is unfamiliar to American readers.

Teachers also provided examples of how they have used the PLR and PLeR to individualize instructional activities. For example, Park Elementary's special education resource teacher talked about a boy in her class who was interested in drawing but had a very limited sight vocabulary. After observing his strength (drawing) and recognizing his weakness (sight vocabulary), she developed activities designed to draw on the strength to help improve his vocabulary and language skills. The teacher and student made books together. The child would draw a story, his teacher would then help him write the story he had drawn, and he would maintain new sight vocabulary words in a "word bank" at home.

Another example was provided by the Park Elementary prekindergarten-kindergarten teacher. She told of how she observed a girl in her classroom playing with blocks. The child was talking to herself, "I'm going to put a red block here and another one here, blue here and here. . . ." Watching her, her teacher recognized that the girl was dealing with issues of symmetry and further realized what types of problems the girl was ready to deal with in math. The teacher also identified ways to support the child's learning through literature (e.g., repetitive motion, as in *The Gingerbread Man*).

Finally, a kindergarten-first grade teacher told how she sometimes uses her observations of children at particular work areas in her classroom to understand not only children's learning better, but also to evaluate the effectiveness of how she is using those work areas. For example, she told of how she felt she had been ignoring the "water table" in her classroom, as it is a very self-sustaining activity (children play with water in a large, shallow "table" and conduct "experiments"). The teacher decided to sit in a corner and take notes as some children worked at the water table. She observed them experimenting with bubbles and recorded the children's "science talk" about "mixing magic potions." The children conducted an inquiry of their own design into why bubbles touched by wet hands didn't break but bubbles touched by dry hands did break. Through her observations of these children, the teacher said she better understood the learning that takes place at the water table and will better be able to guide children in their learning at that work area.

An additional example of how one teacher uses the PLR/PLeR in her classroom appears in Exhibit II.

Evaluation

There has been no formal evaluation of the effects the PLR/PLeR has upon teaching and learning. However, the CCE teacher consultant has completed annual reports documenting the use of PLR during the 1992-93, 1993-94, and 1994-95 school years. The reports describe how the PLR (and in 1994-95, the PLeR) was implemented in classrooms and how the author and other teachers have reacted to its use. She includes comments from all teachers involved at the 10 schools participating in ESAP.

EXHIBIT II

Using the PLeR to Assess Children's Learning

Yvonne Smith's preK-K classroom makes no waste of the modest amount of space available to her 20 or so young students. The classroom has about ten different work areas in which children build with blocks, play at the "water table," draw, paint, and write, dress up and play house, play at the sand table, observe the class' pet gerbils, bake, and read in the "library."

During a two-hour period one morning in April 1995, Yvonne's students worked in pairs or trios at all of these activities. Teacher's aides worked with three pairs who required adult assistance: children baking biscuits for the class to enjoy as a snack later, children drawing pictures in crayon and then ironing their drawings onto fabric, and children sewing dolls of their favorite people. Yvonne moved among the other children in the classroom, observing their activities and talking with them about what they were doing.

Yvonne later talked about the observations she made of children's learning during this period and how she would use the PLeR to record and analyze the observations and to plan future activities for her students.

In one instance, two girls playing in the playhouse called Yvonne over to sing "Happy Birthday" to her. After they were done, Megan, an African-American child who speaks only English, said, "Now we have to sing it in the other language," meaning Spanish (the children in the class, many of whom speak Spanish as their first language, have been learning to sing Happy Birthday in Spanish). Yvonne made a note of Megan's recognition of the fact that there are multiple languages, and she planned to talk more about that fact the next time the children all sang Happy Birthday together as a group.

In another instance, Yvonne talked with a Hispanic boy working at the sand table. She asked him what he had built in the sand, and he said, "a bump." Yvonne knew that this word was a new one for Alberto, whose English skills were still limited. She shared with him two other words he could use to describe what he had built, "hill" and "mountain." She then asked him how many words he had to describe it. He said, "Two." Yvonne held up two fingers and said, "Do you have this many," or, holding up three fingers, "this many?" Alberto realized he had said "two" when he meant "three," and Yvonne realized that he was still translating numbers from Spanish to English in his mind. Yvonne said that she will use what she learned from her observations of Alberto during this period to select reading materials that reinforce his growing English vocabulary (i.e., books that have hills, mountains, and bumps in them) and that use numbers in both English and Spanish.

Those who use the PLR/PLeR stress its validity for individualized instruction. As one teacher put it, "Assessment is instruction," meaning that ongoing assessment of children's progress must guide instruction and that careful, systematic assessment designed to inform instruction, particularly as it is tailored to the individual child, is inherently valid. Teachers at Park Elementary also suggest that the PLeR is an even more valid tool for assessment than the PLR, as, in the words of one teacher, "It feels less cumbersome, and more natural, having more pieces" (i.e., observing all of a child's learning, not just extricating the language learning).

Because teachers use the PLR/PLeR to help them reflect about individual children, where they are in their learning, and how the teacher can best help the child learn, teachers — and the developers of the PLR and PLeR — are not concerned so much about reliability. Insofar as the PLR/PLeR helps teachers make good choices about the instructional techniques and the curriculum they use with their students, the tool, they say, is reliable. Teachers also suggest that interrater reliability is not an issue, except insofar as teachers use it to communicate information about children amongst themselves and with parents. The PLR/PLeR is not used to rate a child's development against others, but to evaluate each child's progress against her or his own baseline performance.[3]

Resource and Training Support

Teachers using the PLR and the PLeR receive and give ongoing training in its use through multiple channels:

- CCE study groups,
- CCE's teacher consultant,
- Continuing Education classes at Lehman College, and
- Park Elementary staff meetings.

The basic structure of support activities has not changed since the introduction of the PLeR.

CCE Study Groups

The CCE coordinates teacher "study groups," which take two forms. One study group, which meets every other week, involves the Park Elementary teachers who use the PLR/PLeR coming together over lunch to discuss their use of the forms and the impact the tool has on their teaching. They share ideas that facilitate their use of the PLR/PLeR. For example, one teacher found it cumbersome to maintain PLR/PLeR records on all 30 children in her class, because the forms were rarely handy when she wanted to record something. She started keeping pads of white stickers with her. When she wants to record an observation, she writes her notes on one of the stickers. At the end of the day, she transfers the stickers onto the observation forms.

More specifically, however, these groups meet to share problems or points of interest they have identified in their classrooms, generally as a result of using the PLR/PLeR. For example, a teacher might share an insight into one of her students, and the group may discuss possible ways of best instructing that student. Frequently, teachers say, by discussing the educational needs of one student, they acquire insights into many others.

---

[3] Though it is not directly relevant to the current case study, a related initiative in California, the California Learning Record, has adapted the PLR/PLeR for use by California educators. The group supporting teachers participating in the work there has performed an analysis of interrater reliability with respect to teachers' assessments of children's reading levels. The group found very high interrater reliability. However, this group had a particular outcome variable they were looking at — the level at which children were reading as designated by a numerical scale — while teachers participating in NYAN do not attempt to identify children's reading (or other) development according to any particular scale. Thus, educators and policymakers who are interested in uniformity of application of PLR/PLeR and interrater reliability of teacher judgments should refer to the work of the California group. Barr, M.A., & Syverson, M. (1994). *The California Learning Record System: Moderation Readings Report*. San Diego: University of California.

In addition to the Park Elementary study groups, approximately once a month CCE brings together teachers from different member schools to share experiences. CCE, Park Elementary, and the developers of the PLR/PLeR all stress that communication among teachers is crucial to improving the craft of teaching. The CCE consultant who works with teachers in study groups calls the sessions "a shot in the arm for teachers." One teacher commented about the collaborative nature of the endeavor: "As teachers begin to do this [talk about their students' learning], they become each other's gurus."

During the 1994-95, two Park Elementary teachers, together with the CCE teacher consultant, led a study group for CCE teachers using the PLeR. According to the CCE teacher consultant, the structure of the study group they led generally included a discussion of an assigned article, reviews of evidence collected by teachers using the PLeR, and a discussion of other issues identified by teachers.

The CCE Teacher Consultant

The CCE teacher consultant who coordinates the teacher study groups often works one-on-one with teachers, visiting their classrooms and sharing her observations. Frequently teachers invite her to observe and comment on a single child they are having a hard time getting to know.

The teacher consultant says that she works one-on-one with teachers primarily at schools other than Park Elementary; teachers less accustomed to observing children closely need her assistance more. The consultant works with a child to learn something about the child's learning style. She then meets with the teacher or group of teachers to discuss her observations. Soon, the teachers are comfortable making the observations themselves: "As you develop a collection of these profiles, the teachers start doing it themselves . . . . I go in to raise teachers' questioning."

Continuing Education Classes for PLR Teachers

Through one of the other member groups of NYAN, the Elementary Teachers Network, teachers using the PLR/PLeR (in conjunction with any of the three groups) may enroll in continuing education classes (e.g., one Park Elementary teacher took a course on the "Reflective Practitioner" during the spring of 1994) at Lehman College. These classes typically meet once every month and are free to teachers working with NYAN member organizations.

Park Elementary's Own Staff Development Activities

Every other week on their own time, Park Elementary teachers have a staff meeting. The meeting is not business-oriented (those meetings occur at other times). Rather, teachers come together to talk about the children in their classrooms. This tradition is a longstanding one at Park Elementary, and it illustrates why and how the PLR came so naturally to Park Elementary's teachers. These meetings resemble the study groups, the only difference being that they are not focused specifically on the use of the PLR/PLeR.

Impact of the PLR and PLeR

Because the PLR and PLeR are consistent with the educational philosophy already espoused by Park Elementary teachers, their impact on the school's teachers and students, as distinct educational tools, remains difficult to evaluate. To those who use it, the PLR/PLeR has helped teachers focus their observations of children and maintain useful records of their observations: the PLR and PLeR have provided a structured approach for what teachers were doing already.

Impact on Teachers

The teachers who have used the PLR and who are using the PLeR have all volunteered to do so, recognizing in the technique something they wanted to do in their own classrooms. They saw the PLR as a promising tool to supplement what they were already doing, and the PLeR proved an even more natural fit. Teachers at the school who do not use the PLR or PLeR are not opposed to their use; they merely did not identify in it enough value to add it to what they were already doing. As one teacher said, "We were doing PLR before we had PLR."

Using the PLR/PLeR requires a significant amount of time and discipline on the part of teachers. One teacher said she reviews and organizes the notes she has taken during the day while traveling home on the subway. This teacher also likes the fact that the PLR and PLeR are what she calls "non-form forms," meaning that she can adapt the system to her own teaching style. Another teacher echoes this sentiment, saying, "They encourage you to make any changes you want to make. That's one of the things I like about it. I play with it. If it isn't useful to me, why use it?"

This teacher also says that, though she does not use the PLeR with all her students, she does use it with several; for these students, she keeps a notebook with several pages devoted to each child. When she observes something she wants to record, she turns to the child's section in the notebook. Still another teacher says she makes sure she observes each child in her classroom at least one time each week. These routines help the teachers ensure that they are making good use of the PLR/PLeR.

Teachers differ with respect to how much they believe using the PLR and PLeR has affected the way in which they work with children. One teacher said that using the PLR allowed her to "fine tune" her teaching. Another, however, said that using the PLR bolstered her confidence in her ability to observe accurately what her students are doing and learning: "I now trust what I see and what I hear." This teacher also values collaborative teaching more than she did formerly, stating, "I know that it's important to share [my observations] with colleagues, parents, and children." Both of these teachers in Spring 1995 said that using the PLR — and now the PLeR — gets easier each year as they become more used to and better at the process.

Finally, teachers also observe that the record keeping they do with the PLR/PLeR assists them in writing the narrative reports that they complete for each child two times a year.

Teacher Response to Substituting the PLeR for the PLR. Teachers at Park Elementary say that they took the switch to the PLeR in stride. Indeed, just as some Park Elementary teachers said that they were "doing PLR" before there was a PLR, teachers who used the PLR also said that they had already begun to use it to track children's progress in all subject areas before they actually started working with the PLeR. Thus, the switch to the PLeR came naturally to them.

Comments from the CCE Teacher Consultant. The CCE teacher consultant who works with teachers using the PLR/PLeR asserts that the experiences of Park Elementary teachers are atypical. Teachers at Park Elementary come to the PLR/PLeR far better equipped to use it than do most of the teachers she works with at other schools. Park Elementary teachers have several factors working in their favor. The school has always valued teachers coming together to talk about children and to figure out ways to meet their needs; consequently, the school has a flexible schedule to allow teachers time to meet, and teachers are willing to put in significant amounts of their own time to work with their colleagues.

Teachers at other schools have, according to the CCE teacher consultant, struggled with the PLR. Whereas the vision of education held by Park Elementary teachers led some of them to choose to use the PLR, and then the PLeR, at some schools it is using the PLR that has led teachers to struggle to alter their vision of educational practices. Says the teacher consultant, "The PLR works in places with administrative support and a vision of the way children and teachers learn. It does not work where there's not a beginning of that vision." She also asserts, "Time and space have to be legislated for teachers to talk to each other," accommodations already in place at Park Elementary when the PLR was first introduced. According to the teacher consultant, some teachers at other schools have dropped out of ESAP because "they can't be seen as good teachers in their schools and look at children the way we do."

The CCE teacher consultant and a Park Elementary teacher who served as a mentor to other teachers using the PLR during the 1993-94 school year spoke to how thoroughly many teachers have embraced the PLR/PLeR. Typical remarks they hear are, "I never really listened to children before. I never really watched them work. I never really paid attention to their choices." The mentor teacher says that some of the teachers she talks to have had to "dramatically change their classrooms" and that they are ". . . now furious with Boards of Education and the institutions that trained them — the beginning point is the child."

Impact on Curriculum and Instruction

As the PLR, and then the PLeR, joined an educational program already very child-centered, their impact on curriculum and instruction at Park Elementary has been mostly at the margin. One teacher says, "Because I have written down what the child is doing, I know what can possibly come next. I know what kinds of materials I need. The PLR helps me individualize instruction more. . . . But there is no huge change [in my instructional methods], only fine tuning." Other teachers at the school also commented that the PLR/PLeR has helped them individualize instruction for their students and identify which educational materials students need. Also, during 1993-94, one teacher of fifth and sixth graders at Park Elementary used the PLeR (prior to its adoption by all ESAP-participating teachers) to help her restructure her mathematics curriculum, integrating it with other subject areas.

Teachers' comments about the impact of the PLR/PLeR in this respect did not change from 1993-94 to 1994-95.

Impact on Students

According to teachers, students take the PLR/PLeR in stride. One teacher reported that her children often ask her, "Are you going to write down what I did?" All teachers reported how using the PLR/PLeR has helped them tailor their lessons to their students' needs; some of their anecdotes have already been included, above.

Impact on Students with Special Needs. One significant use teachers at Park Elementary put the PLR to in 1993-94 was to aid them in their documentation while in the process of declassifying special education students. Park Elementary's special education teacher reported that she used the PLR as part of her documentation of three students' readiness to be declassified as special education students; two of these students were declassified during the school year. This teacher also believes that the PLR/PLeR would make a good substitute for IEPs, which she feels are too rigid and cumbersome; the PLR/PLeR could serve the same purpose but allow more flexibility.

Impact on Parents

Teachers report that parents like the PLR/PLeR, and that they especially like the parent-teacher-student conferences devoted to discussing the child's literacy and other intellectual development; several parents have told teachers they've never been asked before to share their knowledge of their children. Teachers and parents alike are gratified to discover that their observations of their children are so similar. (Note, however, that these parental reactions come filtered through teachers; the one parent we interviewed said that her daughter's teacher does not use the PLR or PLeR. It is, therefore, difficult to determine the impact the tools have had on parents.)

Neither teachers nor the parent interviewed for this study commented on any changes in parental reactions between 1993-94 and 1994-95.

Future Plans

Now that the PLeR has replaced the PLR, teachers at Park Elementary who use the tool no longer have plans to change their use of the technique. They will continue to use the PLeR as it makes sense to them in their classrooms. As a school, Park Elementary has no plans to expand use of the PLeR. Any teacher who wants to use it is welcome to do so, and no teacher is compelled to try it.

Conclusions

Park Elementary is a school in which teacher inquiry and teacher collaboration are encouraged. In such an environment, the interweaving of assessment with curriculum and instruction — as fostered by the PLR, the PLeR, or other techniques — comes naturally to teachers, at least after practice.

The PLR/PLeR is an example of an assessment technique designed, more than anything else, to inform curriculum and instruction. It encourages teachers to hone their observational skills and furthers their engagement in the pedagogical process. It is a tool for teachers to apply in their classrooms to help them work more closely with their students. It is not an assessment technique that aims to provide information about students' progress toward any pre-specified standards of achievement or performance, although, based on their observations, teachers seem to formulate individualized goals for their students. Given various attributes and limitations, it is hard to imagine the PLR or the PLeR being used successfully for purposes of accountability or large-scale assessment of student achievement. Simply put, when teachers commit their time to using the PLR or PLeR to guide their classroom activities, the process most likely is, as the CCE teacher consultant suggests, a "powerful" tool and far "more profound than it sounds."

APPENDIX A

Primary Language Record Form

Primary Language Record

School

School Year

Name

DoB

Summer born child ☐

☐ Boy ☐ Girl

Languages understood

Languages read

Languages spoken

Languages written

Details of any aspects of hearing, vision or coordination affecting the child's language/literacy. Give the source and date of this information.

Names of staff involved with child's language and literacy development.

Part A To be completed during the Autumn Term

A1 Record of discussion between child's parent(s) and class teacher *(Handbook pages 12-13)*

Signed Parent(s) ____________________ Teacher ____________________

Date ____________

A2 Record of language/literacy conference with child *(Handbook pages 14-15)*

Date ____________

© Centre for Language in Primary Education, 1988. Published in the U.S.A. by Heinemann Educational Books, Inc., 361 Hanover St., Portsmouth, NH 03801

Part B

To be completed during the Spring Term and to include information from all teachers currently teaching the child.

Child as a language user (one or more languages)

(Handbook pages 17-18)

Teachers should bear in mind the Authority's Equal Opportunities Policies (race, gender and class) in completing each section of the record and should refer to *Educational Opportunities for All?*, the ILEA report on special educational needs.

B1 Talking and listening

(Handbook pages 19-22)

Please comment on the child's development and use of spoken language in different social and curriculum contexts, in English and/or other community languages: evidence of talk for learning and thinking; range and variety of talk for particular purposes; experience and confidence in talking and listening with different people in different settings.

What experiences and teaching have helped/would help development in this area? Record outcomes of any discussion with head teacher, other staff, or parent(s).

B2 Reading

(Handbook pages 23-28)

Please comment on the child's progress and development as a reader in English and/or other community languages: the stage at which the child is operating (refer to the reading scales on pages 26-27); the range, quantity and variety of reading in all areas of the curriculum; the child's pleasure and involvement in story and reading, alone or with others; the range of strategies used when reading and the child's ability to reflect critically on what is read.

continued

B2 (continued)

What experiences and teaching have helped/would help development in this area? Record outcomes of any discussion with head teacher, other staff, or parent(s).

B3 Writing *(Handbook pages 29-34)*

Please comment on the child's progress and development as a writer in English and/or other community languages: the degree of confidence and independence as a writer; the range, quantity and variety of writing in all areas of the curriculum; the child's pleasure and involvement in writing both narrative and non-narrative, alone and in collaboration with others; the influence of reading on the child's writing; growing understanding of written language, its conventions and spelling.

What experiences and teaching have helped/would help development in this area? Record outcomes of any discussion with head teacher, other staff, or parent(s).

Signature of head teacher and all teachers contributing to this section of the record:

Part C To be completed during the Summer Term*

(Handbook page 35)

C1 Comments on the record by child's parent(s)

C2 Record of language/literacy conference with child

C3 Information for receiving teacher

This section is to ensure that information for the receiving teacher is as up to date as possible. Please comment on changes and development in any aspect of the child's language since Part B was completed.

What experiences and teaching have helped/would help development? Record outcomes of any discussion with head teacher, other staff, or parent(s).

Signed: Parent(s) ______________________ Class Teacher ______________________

Date ______________ Head Teacher ______________________

**To be completed by the Summer half-term for 4th year juniors.*

Observations and Samples (Primary Language Record)

attach extra pages where needed

Name: **Year Group:**

1 Talking & listening: diary of observations

The diary below is for recording examples of the child's developing use of talk for learning and for interacting with others in English and/or other community languages.

Include different kinds of talk (e.g. planning an event, solving a problem, expressing a point of view or feelings, reporting on the results of an investigation, telling a story ...)

Note the child's experience and confidence in handling social dimensions of talk (e.g. initiating a discussion, listening to another contribution, qualifying former ideas, encouraging others ...)

The matrix sets out some possible contexts for observing talk and listening. Observations made in the diary can be plotted on the matrix to record the range of social and curriculum contexts sampled.

(Handbook pages 37-39)

	SOCIAL CONTEXTS				
LEARNING CONTEXTS	pair	small group	child with adult	small/large group with adult	
collaborative reading and writing activities					
play, dramatic play, drama & storying					
environmental studies & historical research					
maths & science investigations					
design, construction, craft & art projects					

Dates	Observations and their contexts

© Centre for Language in Primary Education, 1988. Published in the U.S.A. by Heinemann Educational Books, Inc., 361 Hanover St., Portsmouth, NH 03801

2 Reading and Writing: diary of observations (reading and writing in English and/or other community languages)

(Handbook pages 40-44)

Date

Reading

Record observations of the child's development as a reader (including wider experiences of story) across a range of contexts.

Writing

Record observations of the child's development as a writer (including stories dictated by the child) across a range of contexts.

3 Reading Samples (reading in English and/or other community languages)

to include reading aloud and reading silently

(Handbook pages 45-49)

Dates			
Title or book/text (fiction or information)			
Known/unknown text			
Sampling procedure used: informal assessment/running record/miscue analysis			
Overall impression of the child's reading: • confidence and degree of independence • involvement in the book/text • the way in which the child read the text aloud			
Strategies the child used when reading aloud: • drawing on previous experience to make sense of the book/text * playing at reading * using book language * reading the pictures * focusing on print (directionality, 1:1 correspondence, recognition of certain words) • using semantic/syntactic/grapho-phonic cues • predicting • self-correcting • using several strategies or over-dependent on one			
Child's response to the book/text: • personal response • critical response (understanding, evaluating, appreciating wider meanings)			
What this sample shows about the child's development as a reader. **Experiences/support needed to further development.**			

** Early indicators that the child is moving into reading*

Observations and Samples (Primary Language Record)

attach extra pages where needed

Name: **Year Group:**

1 Talking & listening: diary of observations

The diary below is for recording examples of the child's developing use of talk for learning and for interacting with others in English and/or other community languages.

Include different kinds of talk (e.g. planning an event, solving a problem, expressing a point of view or feelings, reporting on the results of an investigation, telling a story ...)

Note the child's experience and confidence in handling social dimensions of talk (e.g. initiating a discussion, listening to another contribution, qualifying former ideas, encouraging others ...)

The matrix sets out some possible contexts for observing talk and listening. Observations made in the diary can be plotted on the matrix to record the range of social and curriculum contexts sampled.

(Handbook pages 37-39)

	SOCIAL CONTEXTS				
LEARNING CONTEXTS	pair	small group	child with adult	small/large group with adult	
collaborative reading and writing activities					
play, dramatic play, drama & storying					
environmental studies & historical research					
maths & science investigations					
design, construction, craft & art projects					

Dates	Observations and their contexts

© Centre for Language in Primary Education, 1988. Published in the U.S.A. by Heinemann Educational Books, Inc., 361 Hanover St., Portsmouth, NH 03801

Language Arts and Math Portfolios: Niños Bonitos Elementary School

April 18-19, 1994
February 27-28, 1995

LANGUAGE ARTS AND MATH PORTFOLIOS:
NIÑOS BONITOS ELEMENTARY SCHOOL

Introduction

This case study examines a performance assessment system designed at an individual school in response to its own pressures for reform. The subject of the study, Niños Bonitos Elementary School, is located in suburban Niños Bonitos, California, and is part of the San Diego Unified School District (SDUSD). As of the 1994-95 school year, Niños Bonitos Elementary School served a total population of 924 students from pre-kindergarten through sixth grade. Thirty-nine percent of the student population are of Southeast Asian heritage (Vietnamese, Laotian, and Hmong), 46 percent are Hispanic, and 5 percent are African American. Other ethnic groups include East Asians and Filipinos. Almost 95 percent of the student population qualify for free or reduced lunches, and 77 percent have been identified as possessing limited English proficiency (LEP). Eighty-six percent qualify for Chapter 1 services (statewide, only 18 percent of the student population are eligible for Chapter 1 services). The five major languages spoken by students are Vietnamese, Hmong, Lao, Spanish, and English.

Niños Bonitos is located in a suburban community close to the University of San Diego. The neighborhood immediately surrounding the school served as a relocation site for Lao and Hmong refugees arriving at nearby Camp Pendleton in the mid-1970s. It is an area that still struggles with poverty and its related problems. As an indication of the poverty faced by many of Niños Bonitos' children and families, the school is seriously considering purchasing uniforms for all students to wear, since so many families cannot afford to clothe their children adequately.

Although the study focuses on reforms undertaken at the local level, it first includes some discussion of district and state reforms that interact with and have an impact upon local initiatives.

Participants

At Niños Bonitos, the people identified in Exhibit I were interviewed.

Observations

In addition to conducting interviews with the listed participants, during the 1993-94 school year the study staff observed a four-hour staff development session that included both a presentation by the mathematics curriculum committee responsible for developing the school's mathematics performance assessment system and a video explaining California's impending state-wide performance assessment system. In 1994-95, the study staff observed a third-fourth grade (middle wing), transitional classroom for four hours.

EXHIBIT I

Study Participants

1993-94	1994-95
- District assessment specialist - School principal[*] - School vice principal - A school resource specialist - The school's teacher union representative[*] - One mathematics teacher (early childhood wing) - One parent who also serves as the school's parent volunteer coordinator	- District assessment specialist - One school board member - School principal[*] - A second school resource specialist - The school's teacher union representative[*] - Two language arts teachers (early childhood and upper wings) - One mathematics teacher (upper wing) - One parent

District and State Context

In 1994-95, San Diego Unified School District included 109 elementary schools, 21 middle schools, 16 high schools, and 7 "atypical" schools, all together serving approximately 130,000 students. Approximately 17 percent of the SDUSD student body is African American, 18 percent is Asian, 31 percent is Caucasian, and 33 percent is Hispanic. Close to 60 percent of the students qualify for Chapter 1 funds, and about 70 percent receive free or reduced lunch.

Under both its former and its new superintendent, the SDUSD has been a major advocate of innovation and reform, and it has taken a particular interest in developing and implementing alternative assessments of student performance. In 1991, Assembly Member Quackenbush introduced AB 40, a local piece of legislation designed to fund research, development, and dissemination of alternative assessments. The California Assessment Collaborative (CAC), composed of schools, districts, and county offices in northern California and the San Diego Consortium (consisting of San Diego City and San Diego County schools), was invested with the responsibility for carrying out these three functions.

Under the San Diego Consortium, the SDUSD became involved in several assessment projects, with teachers taking the lead in developing the actual assessment tools. Those projects include:

- Chapter 1 Portfolio Assessment Development Across Levels;
- Dance Portfolio Assessment Project;

---

[*] This position changed hands between the first and second years of study.

- K-4 Learning Assessment Project;
- Performance Assessment Collaboratives in Education (PACE) Portfolio Project;
- Standards and Assessments for English Learners; and
- The University of California at Los Angeles Center for Research in Educational Standards and Student Testing (CRESST) U.S. History Project.

According to the district assessment specialist, although all of these projects generated a great deal of enthusiasm and interest district educators questioned the wisdom of participating in numerous, isolated projects, rather than building a unified district capacity to confront the task at hand. Hence, the district became linked with several major national movements to promote performance assessments, most notably the New Standards Project (NSP)[2] and the National Alliance for Restructuring Education, part of the New American Schools Development Corporation (NASDEC).[3] Although these affiliations do not provide the district with large financial resources, they do provide access to much-needed expertise and models for educational reform, particularly in the areas of curriculum, instruction, and assessment.

At approximately the same time as SDUSD became allied with these national partners, its own district-sponsored Accountability Committee of educators, parents, and citizens launched a "homegrown" network of 10 (and later 25) schools, which volunteered to serve as models in a Leadership in Accountability Demonstration (LAD) project. The project focused on increasing school accountability through the mechanisms of improved standards, new forms of assessment, and improved public reporting. Although all of San Diego's 160 schools are eligible to benefit from National Alliance and NSP expertise, these 25 LAD schools are the leaders in testing and implementing National Alliance and NSP reforms and will serve as mentors to other district schools through a cluster feeder pattern.

The National Alliance has promoted district adoption of five "design tasks" for improving student achievement and organizational effectiveness. The first design task deals with curriculum, instruction, assessment, and technology. SDUSD's new superintendent has structured her district-wide reform efforts around these five design tasks, and over the past year she developed 16 *expectations* or outcome measurements to test, objectively, whether or not progress is being made in each of the five areas.

Unfortunately, one of the most critical instruments for measuring student achievement — results on the new California Learning Assessment System (CLAS) state test — is no longer available. The CLAS was initiated by the state legislature in 1991 as a new *authentic* assessment of higher-order thinking in the major academic subjects and was piloted in California schools in language arts and mathematics in 1992-93 and 1993-94. The state test was to consist of open-ended and enhanced multiple choice problems that students would work on in groups and then individually respond to in writing. Students would not only have to solve the problems, but would have to explain or justify their solutions as well. However, California's governor vetoed the legislation last summer as a result of pressure from conservative groups and special education groups who were concerned about the assessment's content, method of administration, and equity.

---

[2] California is one of 19 NSP state partners, and San Diego is one of six urban school district partners.

[3] San Diego is one of 9 National Alliance sites, all of which are also NSP partners.

The district's assessment specialist, as well as the school board member interviewed, noted that the district has been left somewhat "rudderless" by the demise of the CLAS, since CLAS results were to be the basis for measuring progress in meeting the National Alliance design tasks and for measuring success among the LAD schools. The school board member said that school officials felt quite strongly that the CLAS would be a better measure of school improvement than what has traditionally been used — the standardized Abbreviated Stanford Achievement Test (ASAT.) The superintendent is still a strong proponent of the use of performance assessments within the district; within a week of the CLAS veto she had issued a report to the school board regarding the status of performance assessment in the district, advocating increased use of NSP assessments by all schools in the district as a replacement for the CLAS. In the spring of 1995, district staff conferred with NSP personnel about using the NSP standards and assessment strategies in every school in the district, thus making SDUSD as a special "action research laboratory" for the NSP program.

School Context

Considered a leader in assessment reform, Niños Bonitos Elementary School is a LAD school, and its Principal (the Principal during the round one visit) was at one time on the LAD committee. In addition, Niños Bonitos is one of eight "1274" schools in the district schools that are part of a legislatively mandated state network; network comprised of schools that are leaders in reform and restructuring efforts. Through its affiliation with LAD and the 1274 program, Niños Bonitos draws heavily upon the National Alliance and NSP approaches and is continually asked to share its assessment practices with the rest of the district. Because Niños Bonitos' history of reform places it in the vanguard of SDUSD schools, the school is, at this point, giving more to the district than it is getting in the way of technical assistance and expertise.

The school occupies two separate sites and operates on a year-round schedule, employing two administrators, 30 classroom teachers, and 52 instructional aides. In addition, the school maintains a full complement of support staff, including resource teachers, a librarian, a counselor, a nurse, a volunteer coordinator, community aides, and custodial staff. Although nearly half of the staff speak a second and often a third language, only two credentialed classroom teachers speak any Southeast Asian languages (one speaks Vietnamese and the other speaks Hmong).

Niños Bonitos' elementary school is democratically governed through a staff committee structure coordinated by the school principal. All certified and classified staff must serve on both a site committee (e.g., student involvement, grants, finance, personnel) and a curriculum committee. Committees make all major decisions related to the school's operations, although committee recommendations are always submitted to the full staff for a consensus review.

Niños Bonitos' Principal in 1993-94 was described by the teachers included in the study as an "excellent leader." She initiated the school's comprehensive restructuring process in 1988, and recently was awarded the honor of "California Principal of the Year." At the end of the first year of this study, she was chosen in a national competition for the position of *Principal in Residence* with the U.S. Department of Education. As a result, the school acquired a new Principal in 1994-95, the second year of the study.

History of School Reform

The development of the school's performance assessment has been heavily influenced by the school's demographic character. In 1988, after six months of careful observation, the school principal determined that children's needs — particularly for learning the English language — were not being met appropriately. Niños Bonitos' ASAT scores[4] were extremely low, and staff morale was quite poor. She felt more appropriate instructional programs and language proficiency groupings were needed to improve learning and to provide greater equity of services to Niños Bonitos' diverse student population. In addition, as an early proponent of site-based management and accountability, she felt the full staff should take responsibility for developing the mechanisms of change.

To initiate the restructuring process, the Principal developed a site-based committee structure of governance for the school, which proceeded to revamp the school curriculum and to reorganize the school day. Emphasis was placed initially upon language arts, historically the area of students' greatest weakness and of staff's greatest frustration. Students were organized into four nongraded but age-appropriate "wings." At the school's Annex, pre-kindergarten, kindergarten, and first graders who are severely handicapped or in need of developmental services would participate in a special early childhood education program. The primary (grades 1 and 2), middle (grades 3 and 4), and upper (grades 5 and 6) wings were housed at Niños Bonitos' main campus.

Niños Bonitos new organizational system worked very well and was adopted permanently in 1989. In the morning, all students in each wing (except those who are Spanish-speaking) are assigned to one of six levels of English proficiency (e.g., entry, Sheltered A, Sheltered B, Transition A, Transition B, and non-sheltered) for language arts and mathematics coursework. For example, a first grade student with limited English proficiency might be assigned to a primary wing Sheltered A classroom, while a 5th grade student with near fluency in English might be assigned to an upper wing Transition B classroom. Spanish-speaking children are enrolled in a bilingual program, in which all three basic subjects are taught in Spanish, supplemented by additional English-language development. In the afternoon, a reading recovery course is offered to first graders who are experiencing difficulty with beginning reading.

In the early stages of her restructuring efforts at Niños Bonitos, the school's principal made sure to discuss her plans with district personnel and to obtain their "buy-in." As a result, her efforts encountered almost no resistance from the district and, in fact, received a great deal of focused attention and support. Her job was made easier by the fact that the school's superintendent and staff were themselves struggling to conceptualize and implement a number of important educational reforms.

---

[4] Each year, as a "Chapter 1 school," Niños Bonitos administers this test to all English-speaking students in reading comprehension and mathematics. Spanish-speaking students take a Spanish equivalent of the test, called the "Aprenda". "Sheltered" students speaking primary languages other than English take the ASAT only after attending Niños Bonitos for one year.

The school's new principal in 1994-95 is extremely impressed by what he has "inherited" at Niños Bonitos. He believes the key to the school's success is empowering the whole faculty to serve as instructional leaders and to experiment with "cutting edge" instructional and assessment reforms.

Performance Assessment Development and Characteristics

Curriculum and assessment reform at Niños Bonitos also began with the language arts program. Staff felt it was important to establish clear entry and exit criteria for every child at each age and language development level as well as clear instructional goals for each level. They wanted to be able to continuously monitor the progress of each child so that an LEP review team could move children to new language development levels at any time during the school year, as those children became ready to advance.

Language Arts

To meet their objectives, faculty "wings" began to meet in 1988 to brainstorm the development of language arts learning outcomes and scoring rubrics — a process that ultimately took four years to complete. Learning outcomes, instructional materials, and portfolio contents were developed for each language development level within each wing (e.g., Primary Level/Entry, Primary Level/Sheltered A). For each level, the learning outcomes describe specifically what students should know and what they should be able to do with regard to oral language, reading, and writing in order to complete a particular level.

If teachers keep these learning outcomes in mind, assessment becomes an integral, daily part of instruction. For example, an entry level LEP student at the primary level should be able, among other things, to use proper words to identify objects through oral language. In reading, that same student should be able to answer simple questions regarding the structure and meaning of a story.

With the help of experts such as Dennie Wolf and Grant Wiggins, Niños Bonitos teachers agreed that portfolios of student writing, along with observations on students' reading and spoken language collected throughout the year would be another exemplary way to measure growth in language arts. Teachers then spelled out what each portfolio should contain for each age and language development level. For instance, the primary level portfolios for Spanish-speaking students should contain three daily journal samples, a minimum of three independent writing samples, and a reading and oral language usage checklists.[5] (The portfolios then are scanned to generate electronic portfolios).

In the second year of our the study (Round 2), teachers began to reevaluate the way in which the language arts electronic portfolios are constructed. To date, teachers have selected the student pieces to be included in each portfolio. However, they are beginning to feel that students should have a greater hand in the selection process, so that the portfolio becomes a tool for student self-reflection, as well as a method of measuring student progress from the teacher's perspective. This year, for instance, the middle wing teacher observed in the study allowed her students to choose their "writing process piece" for their electronic portfolios after the first grading period. The observation is described below in Exhibit II.

---

[5] Reading and oral language checklists were developed as an observational tool, in order to measure student growth in these areas three times a year.

Scoring rubrics were developed by teachers as a way to compare and communicate students' academic progress without having to "label" children with standard letter grades. Language arts scoring rubrics were modified and refined between the two years of the study; by 1994-95, the faculty had identified five performance levels for oral language and reading, and six levels for written language. These rubrics cut across language development and age levels. In writing, for instance, a student can be evaluated as a *pre-writer*, *emergent*, *developing*, *beginning*, *experienced*, or *exceptionally experienced writer*. Specific (and different) criteria were determined by age and performance level. Appendix A provides a copy of Niños Bonitos'language arts scoring rubric. The rubrics themselves were used to redesign the student in language arts, which was renamed the "student growth record" (see Appendix B). Students are now assigned a rubric score by their language arts teacher based on both a review of a written language portfolio and an assessment of performance on reading and oral language.

EXHIBIT II

Classroom Observation

In the 3rd/4th-grade ("middle wing") classroom observed, "transitional" students (those who are almost, but not quite, fluent in the English language), spent most of the morning working in cooperative groups of four, using six personal and laptop classroom computers. Their teacher had received five days of special training in software technology, as one of the school's five "Apple Classroom of Tomorrow" (ACOT) teachers. As a result of her training, she had designed a performance task that required students to describe and illustrate a book they had read about the difficulties Southeast Asian students (like themselves) experience as they assimilate into their new American culture. She adopted the role of "coach," as specified in her training, and circulated throughout the classroom as her students worked — helping them with their writing as well as with their software skills.

The teacher expressed great enthusiasm for the "student-centered" approach she had learned through the ACOT program and she had developed a number of creative ways to integrate technology into her language arts teaching and assessment methods (for example, earlier in the year she had asked students to write haiku poems to appear on T-shirts they designed on their computer screens). Students saved their work in both electronic and "hard-copy" versions of their language arts portfolios, which were shared with parents at parent-teacher conferences, which occur three times a year.

Teachers interviewed noted that the language arts rubrics will most likely undergo continued refinement as the faculty's experience with them and exposure to other models increases. In the second year of the study, for instance, teachers were discussing refinements that would make their rubrics more consistent with the NSP standards.

Mathematics

Language arts standards and rubrics have been in use for four academic years, and their electronic versions for three. In 1993-94, Niños Bonitos staff began to research and write mathematics standards and rubrics, using the same process that they had used for language arts. The newly developed California state frameworks and curriculum guidelines and the National Council of Teachers of Mathematics (NCTM) standards were used as guides to identify eight critical content areas of the mathematics curriculum that students need to master, including: number sense and numeration, whole number operations, whole number computation, geometry and spatial sense, measurement, statistics and probability, patterns and relations, and fractions and decimals. Using NCTM materials, staff also identified five process skills necessary to understand the subject (e.g., problem solving, communications, reasoning, connections, and estimation).

Students in each wing complete 22-day mathematics instructional rotations organized around the NCTM content areas. For instance, the middle wing transitional class observed in 1994-95 begins the year with a division and multiplication rotation taught by their own homeroom teacher, and then, after 22 days, the class rotates to a measurements unit taught by a second teacher, and so on, so that they complete seven content area rotations by the end of the year. Each teacher, then, teaches his or her "specialty" rotation seven times over the course of the year to seven different groups of students.

From the process and content skills they developed, the mathematics committee generated a scoring matrix that was used in place of the current mathematics portion of the growth record to evaluate student progress for the first time in 1994-95. Appendix C presents a copy of this matrix. Students are identified as either *early emergent, emergent, developing, moderately experienced, experienced,* or *exceptional mathematicians* within the seven critical content areas identified above.

For the early childhood wing, the committee also developed a list of activities and projects for each level — the idea being that teachers could use these to assess student mastery of each process skill within each content area. For example, to demonstrate problem-solving skills for the *patterning content* area, kindergartners must show how they can continue the formation of a pattern of paper apples the teacher has begun. To demonstrate pattern reasoning, they must be able to reply verbally to the question, "What comes next"?

After considering mathematics rubrics, the entire faculty, meeting in groups by wings, went on to develop *learner outcomes* for each age level — outcomes that conceptualized what students should know in order to demonstrate mastery of each content and process skill area. The mathematics curriculum committee then developed separate *observable student behaviors* to accompany each set of learner outcomes. These observable behaviors specifically detail tasks each student must be able to perform before moving to the next level.

In 1994-95, these learner outcomes and observable student behaviors were being used to develop *High Expectations Learning Plans for Students* (HELPS) study units recommended by the National Alliance for Restructuring Education. HELPS units require teachers to develop specific activities, student products, and assessment tools for each NCTM concept area (such as geometry and spatial sense or fractions and ratios — these match the 22-day instructional rotations in mathematics at Niños Bonitos). For each activity, teachers must specify the knowledge they expect students to gain (the "learner outcomes") from the activity and the skills they expect students to demonstrate (the "observable student behaviors"). Each instructional rotation ends with a "culminating activity" that requires students to employ most of the knowledge and skills they have learned. For instance, in the middle wing classroom observed, students created three-dimensional castles built of basic geometric shapes as a culminating activity. Students had to provide formulas for the perimeter, area, and volume of their castles in a computerized report and prepare an oral presentation on their project. Teachers interviewed commented on how much they felt they were gaining from the HELPS approach this year, and how well they felt the HELPS philosophy matched their own "homegrown" attempts at reform in mathematics.

At the time of the first visit a mathematics portfolio was being planned, and by 1994-95 mathematics portfolios had been introduced at all grade levels, although in differing forms and at differing paces. The movement toward portfolios is led by two upper wing teachers who are piloting NSP portfolios and sharing their knowledge about portfolio procedures with their colleagues at Wednesday in-service meetings. Ideally, teachers are to follow the NSP model for portfolios, which asks students to choose works demonstrating understanding of each major mathematical content area and to justify their choices through letters of introduction to the reader.

In the upper wing, students are asked to choose a "favorite problem or project" from each rotation and to write about why they enjoyed it and what they learned from it. In the middle wing classroom observed, students are asked to write periodic entries in a "mathematics journal" for the same purpose. Some teachers, however, are using the mathematics portfolios simply as a holding file for all student work, making little attempt to select exemplary pieces or to ask students to reflect upon them. The middle wing teacher said she uses her students' portfolios for two purposes: to support her choice of scoring rubric for each student and to provide concrete evidence of student progress in parent-teacher conferences.

Assessment Quality and Consequences

During the 1994-95 school year, no formal evaluations of Niños Bonitos' performance evaluation system have been undertaken. However, the school's staff continues to express the conviction that their system of measuring student performance has more validity than traditional kinds of evaluation. Staff members maintain that the portfolios and assessment tasks they have enumerated provide more complete information about the exact nature of student competency. They also maintain that their locally designed assessment mechanisms would have enhanced student performance on the CLAS. Unfortunately, as the new principal pointed out (and as noted earlier), the ASAT is likely a poor measure of student achievement at Niños Bonitos, since it tests fact-based knowledge and skills rather than students' understanding of concepts and problem-solving abilities. Nonetheless, Niños Bonitos' reforms seem to have had a salutary effect on student ASAT scores, since in the second year of the study Niños Bonitos was one of three schools in the district to receive state-wide recognition as a "Chapter 1 school," based on its relatively strong ASAT scores.

In order to build greater *inter-rater* reliability into their assessment system in language arts, the teaching staff last year developed portfolio *anchor papers* to support each rubric. Those papers provide a staff-approved exemplar of student work for each rubric at each language development and grade level. In addition, teachers note that the rubrics are so specific as to leave little room for major judgmental differences. At their regular staff meetings, teachers have held informal "scoring conferences" to test this theory and have found high levels of agreement among readers about language arts scores student work should receive. In 1994-95 the new principal, too, was enthusiastic about the reliability of rubric scoring, noting that he raised questions about less than 10 percent of scores after reviewing the student portfolios to which they referred. However, as of 1994-95 no anchor papers had yet been developed for mathematics, although teachers engage in informal consultations about appropriate scores for student work. One teacher noted that she will not have complete confidence in the reliability — and hence, fairness — of her mathematics rubric assignments until mathematics anchor papers are developed — perhaps as early as the 1995-96 school year.

The San Diego Unified School District has demonstrated its faith in Niños Bonitos' assessment system by authorizing the use of language arts rubrics in a narrative report card that substitutes for the standard grading system. Furthermore, by the second year of the study the school had received a waiver to use the new mathematics assessment matrix as a substitute for the mathematics portion of the standard district report card.

In addition to reflecting student work, student *improvement* on portfolio writing samples completed throughout the year are used as one criterion for judging staff performance (along with a review of student growth records, checklists, classroom observations, and assessments of teacher participation on governance committees). Ironically, California prohibits school administrators from evaluating teachers on the basis of standardized test results, but there is no such prohibition against the use of alternative assessments as evaluation criteria. In support of this use of alternative criteria, the Vice Principal maintains that she can identify signs of poor teaching more readily through a review of student portfolios than she can from a review of student grades. In 1994-95 the new principal, too, was enthusiastic about the use of portfolios to measure the actual student progress that is taking place in each classroom.

Several teachers noted that rubrics have some drawbacks; for instance, students often drop back several rubric levels when moving from one wing to another, which (erroneously) suggests to students regression rather than progression. Secondly, it is often very difficult for non-English-speaking students to attain the highest levels at any wing, although they may achieve tremendous growth *within* a rubric level. One teacher noted that portfolios are helpful in demonstrating this kind of "within rubric" growth to both parents and students.

Resource and Training Support

As mentioned previously, Niños Bonitos' principal initiated the restructuring effort that led to the development of the school's performance assessment system, and educational experts were not called in until the process was well underway. Teachers noted that Niños Bonitos was the first school in the district to undergo significant curriculum and assessment reform, so that, initially, not much help was available, particularly in language arts. As a result, teachers said that they always have received the greatest assistance and support from their own colleagues at Niños Bonitos. The district, however, has provided a great deal of encouragement and some regulatory flexibility, especially in recent years, to support the school's continuing reform efforts. The County Office of Education has also provided some technical assistance, particularly when the district cannot provide the necessary staff development expertise. Both the district and the county took a great interest in Niños Bonitos' efforts because they felt that the success of the CLAS system would soon propel other San Diego schools in the direction of performance-based assessments. Although the CLAS system was dismantled after the first year of study, SDUSD is still heavily committed to the notion of curriculum and assessment reform.

In 1990, the RJR Nabisco Foundation awarded Niños Bonitos Elementary a three-year Next Century School grant of $550,000 to improve and expand the school's reform process. This and an innovative technology grant from the state of California awarded in 1992 have allowed the school to perfect the language arts portfolio process and to record portfolios electronically to accompany students' growth records.

The teachers interviewed, including the school's union representative, felt the release time and training they received were quite adequate to support implementation of the school's new assessment system. Every other Wednesday afternoon is used for planning, for completing students' growth records, and for recording portfolios electronically. In addition, teachers have managed to reserve three 45-minute preparation periods for themselves by cooperating in teaching one another's classes. Teachers noted that fear of technology made them initially reluctant to use computers to electronically record student portfolios, but that the "crisis period of change" has passed.

Because budget constraints do not allow the district to provide extra financial support for Niños Bonitos' efforts, the school's grants committee has successfully sought outside monetary assistance. As a National Alliance school, Niños Bonitos received an Apple Classrooms of Tomorrow (ACOT) grant that has supplied four large hardware systems and ongoing teacher training for four upper wing and one middle wing teacher to support the maintenance of electronic portfolios and the use of technology in classrooms. The National Alliance affiliation also provides other resources, including staff development opportunities. Two teachers who sit on the mathematics curriculum committee, for example, were sent to Harvard's PACE summer mathematics training program. Others have attended the major National Alliance conferences that have been held jointly with NSP conferences each year at different sites across the country.

Interaction with Other Assessments

Teachers pointed out that six years ago, when they began their language arts reforms at Niños Bonitos, very little was available in the way of national, state, or local models to inform their efforts. However, in mathematics, they feel they have not had to start as completely "from scratch," because more national and local expertise is available to guide them. For instance, by the time Niños Bonitos became an NSP partner school, language arts rubrics and portfolios were already in place. On the other hand, NSP, NCTM, and National Alliance materials have assisted teachers a great deal in developing mathematics rubrics, portfolios, classroom projects and activities, and authentic assessments. This year, particularly, Niños Bonitos' teachers have begun to use NSP materials as a "mirror" through which to evaluate their own efforts against those of a highly respected national movement.

At the time of the first visit, Niños Bonitos planned to administer the new performance-based CLAS assessment in 1994-1995 — in mathematics and language arts at grade 4 and in social studies and science at grade 5.

The 1993 CLAS mathematics pilot assessment results from Niños Bonitos showed that none of the 54 students whose assessments were scored were rated above the mid-category of the

6-category performance level scale, with more than half being classified at the lowest performance level. On the other hand, in reading-literature, of the 45 students with scored assessments, 2 percent were classified in the second highest performance level (5th level) and only 9 percent were classified in the bottom performance level.

The CLAS assessment was to be scored by rubric, just as is student performance in Niños Bonitos' language arts and mathematics curriculum. Like SDUSD's district assessment specialist, Niños Bonitos' administrators said that they and other teachers were relatively pleased with the CLAS, since it matched the school's philosophy of providing authentic assessments and grading criteria to drive improved instruction in the classroom. They believe that Niños Bonitos' reforms will continue to bear fruit in the form of improved student ASAT and Aprenda scores, but they are concerned that the these standardized tests do not accurately measure the what students know and are able to do as well as do performance-based examinations like the CLAS.

Impact of Niños Bonitos Performance Assessments

The impact of performance assessments on the Niños Bonitos Community is discussed below.

Impact on Teachers

Although Niños Bonitos' new curriculum and assessment system was primarily focused on student benefits, the reform effort was also intended to boost "what was perceived at the school to be low staff morale." Niños Bonitos' teachers were provided with a number of incentives to adopt the new system, not the least of which was their *full ownership* of its development and implementation. In addition, the new configuration of classes reduced individual class size and provided teachers with more weekly preparation time (teachers only teach one class during the afternoon rotation period). Some of the numerous grants the school has received also allow teachers more release time and more funds to participate in staff development programs. Throughout the districts, six to eight days are reserved for staff development. During those days, the full Niños Bonitos staff meet to work on projects such as the development of standards, rubrics, and portfolio assessments. In addition, every other Wednesday is a *minimum attendance day* for students at Niños Bonitos, in which committee work, including assessment work, is completed.

The Principal noted that the new system has dramatically improved classroom instruction. Most teachers have moved away from the traditional teacher-centered approach to instruction and provide, instead, more cooperative and experiential learning opportunities that focus on problem-solving. One teacher noted that these new approaches, such as portfolios, greatly improve her analysis of students' real progress and achievement levels.

New instructional and assessment methods have also, as previously mentioned, reduced discipline problems and provided teachers with a greater sense of accomplishment, which is augmented by the knowledge that they are personally responsible for the new system's success. In addition, according to their union representative, teachers feel the portfolios make communication with parents much easier. Niños Bonitos' annual staff surveys demonstrate that, overall, teachers are very pleased with the new system and its results.

The overall "costs" of the system are borne, unfortunately perhaps, exclusively by Niños Bonitos staff. Development of the rubrics and the portfolio system required a tremendous amount of work and coordination, although staff were allocated a fair amount of staff development time to accomplish their purpose. In addition, portfolio scoring and rubrics assignment is generally a very time-consuming process. Also, it is apparent that there was a great deal of resistance to the implementation of the system to record portfolios electronically, since staff were required to both learn a great deal of software and to find time outside of class to scan their students' work into the computer system. This latter problem has been somewhat alleviated by asking classroom aides and computer support staff to scan the work; complaints also have decreased as teachers have become "more comfortable with the system."

The Vice Principal noted that performance assessments place additional burdens on teachers since they often must meet the demands of both old and new systems. For example, in the first year of the study, teachers had to complete both the new mathematics assessment scoring matrix and the mathematics portion of the traditional report card (until the district approved the exclusive use of the matrix).

The 1993-94 principal noted that although the local teacher's union pays "lip service" to the need for instructional and assessment reform, it has in fact presented the greatest barrier to its implementation, since the union allows the Principal no flexibility in hiring teachers who are committed to or trained in the use of the new system.

Impact on Students

Teachers and administrators included in the study feel that curricular and assessment changes have had "a palpable effect" on students. Both teachers and parents confirmed, for example, that children are completing fewer worksheets and using more *manipulatives* in their courses. In addition, both teachers and students commented upon the large amount of writing students were doing, both in language arts and in mathematics. One teacher noted that although new methods may slow the pace of instruction a bit, they reinforce the fundamentals and foundations of mathematics and language arts, thereby promoting academic success in the long term.

The teaching methods that the use of rubrics and portfolios support provide students with a sense of purpose, accomplishment, and enjoyment. As a result, teachers experience far fewer discipline problems (only one or two suspensions a year) and a much greater degree of order in their classrooms. The principal has personally witnessed an increase in student engagement and an improvement of "on-task behavior" in her recent classroom observations. Furthermore, student attendance rates have improved dramatically since 1988 and, at 96 percent, attendance at Niños Bonitos is now ranked the best in the district.

The parent included in the study noted that portfolios allow children to see more easily their own improvement. That parent feels, therefore, that students are motivated to strive harder than they did without the portfolio system. Moreover, she feels that weaker students do not experience the same sense of failure they might with standard letter grades. The vice-principal pointed out that the portfolio system "builds greater interaction" between the teacher and child as well as a greater sense of mutual respect, particularly for LEP students from other cultures who find it quite a struggle to express themselves in the traditional classroom format. Almost every teacher interviewed commented upon the improvements in self-confidence and motivation they have seen among students since the reforms began.

These improvements in student attitude and motivation can be objectively measured, and are reflected in student performance, particularly in the areas of oral and written language development for LEP students. Anecdotal evidence further suggests that Niños Bonitos graduates, particularly Chapter 1 students, are performing much better at Montgomery Junior High School, than they used to. Improvement in ASAT scores, however, was spotty between the two years of the study — at some grade levels, scores were up substantially in both reading and mathematics, while at other grade levels, scores were virtually unchanged or down.

In 1994-95 the new principal prepared a detailed analysis of ASAT scores by grade level, recording strengths and weaknesses in each subject area. He hopes teachers will address these strenghts and weaknesses in their classrooms. For instance, in reading comprehension in the 1st grade, students were strong in "predicting outcomes," "drawing conclusions," and "matching."

A parent said that her former Niños Bonitos child was more advanced than his peers in both language arts and mathematics at his new junior high school. Finally, on last year's pilot CLAS administration, Niños Bonitos students scored much higher in language arts than the state average (but much lower than the state average in mathematics, a fact that teachers feel supports both the success of their language arts program and the need to revise their mathematics curriculum).

Impact on Parents

Although many of Niños Bonitos' parents are not literate, they are, according to teachers and administrators, very supportive of the education their children receive under the new system. Because parents at Niños Bonitos are culturally and linguistically diverse, the school makes special efforts to involve them, providing monthly parent outreach evenings and ensuring that translators are present at parent-teacher conferences whenever needed. These interactions with parents have introduced them to the reforms at Niños Bonitos while they are occurring. This year, for instance, one parent outreach evening centered around the topic of mathematics portfolios and another covered language arts portfolios.

The parent interviewed in 1993-94 (also the school's parent and volunteer coordinator) said that parent reaction to the portfolio and rubric system is extremely positive, as measured by both the annual parent surveys during 1992 and 1993 and the feedback received at parent outreach meetings. The parent interviewed in the second year of the study reiterated this opinion. Both parents noted that the portfolios give parents a much more visual and concrete sense of their child's progress, particularly when they have difficulty reading the English language. They added that although some parents initially miss the letter grade system, most adapt to the new rubrics quite easily and eagerly (although some teachers felt that the rubrics are too complicated for non-English-speaking parents to understand). The parents said that the rubrics and portfolios provided them with much more guidance about the ways they can help their children at home with language arts development. Currently, both language arts and mathematics portfolios are used in triennial parent-teacher conferences to provide concrete demonstrations of student growth.

Future Plans

Niños Bonitos' 1993-94 principal said that her goal for the future was to continue to drive changes in instruction through changes in assessment, hopefully in new areas of the curriculum, including science. She desires to continue to convince parents and the general public that portfolios are valid indicators of success and to involve those parents and the public in the developmental process itself.

In 1994-95, Niños Bonitos' new principal conducted a "transition" workshop with the school's teachers to determine what the continuing focus of the school's reform efforts should be and how fast the pace of reform should continue to proceed. The principal said that he learned through this process that Niños Bonitos' teachers were "spread very thin" and were not eager to take on additional new reforms this year. As a result, the principal's efforts this year have been directed at consolidating and strengthening existing innovations, such as refinement of the language arts rubrics and portfolio system and development of mathematics anchor papers to support the new mathematics rubrics. Both he and Niños Bonitos' teachers eventually hope to adapt their system of performance assessments and scoring rubrics to the discipline of science — perhaps as early as next year.

The school board member interviewed in 1994-95 said that, surprisingly, there has been little district political opposition to performance assessments and other educational reforms; she anticipates that, with the continued strong support of the new superintendent, the district as a whole will steadily expand the use of performance assessments.

Conclusions

Niños Bonitos' student population is unusual in that the vast majority of its students are both economically disadvantaged and limited English-language speakers. Centered around written and oral portfolios and specific learning outcome tasks, Niños Bonitos' performance assessment system has been designed to address the special needs of its particular population. Performance assessments have been focused initially on language arts and are used to monitor student reading, writing, and oral language development. Mathematics performance assessments are in development — and similar systems in other disciplines are being contemplated.

As of this writing, no objective data has been gathered concerning the validity or reliability of Niños Bonitos' performance assessments, but staff have received outside "validation" by consulting assessment experts on a regular basis and by developing procedures that foster reliability (such as anchor papers).

Although no formal evaluation of Niños Bonitos' assessment system has been performed, anecdotal evidence from staff and parents suggests that the system's benefits outweigh its costs in terms of improved staff morale, better instructional practices, a more open classroom atmosphere, and increased student motivation. At least in language arts, this evidence is supported by objective measures of student performance that show encouraging progress.

Overall, the future of Niños Bonitos' performance assessment system appears bright, since the school possesses strong leadership and has received support and accolades from the San Diego Unified School District.

APPENDIX A

Early Childhood Rubrics

EARLY CHILDHOOD RUBRICS LISTINGS

Revision 8/13/92, 5/28/93, 10/93, 9/21/94

ORAL LANGUAGE

Silent/Emergent Listener-Speaker
- does not yet respond verbally
- minimal evidence of listening/speaking skills, uses gestures and simple words

Limited Listener-Speaker
- one word response, naming items in picture
- limited vocabulary
- rarely contributes in group settings

Developing Listener-Speaker
- describes pictures using phrases, two-three word responses and/or simple sentences
- occasionally contributes in small group settings

Capable Listener-Speaker
- uses complete sentences that tell about a picture
- occasionally contributes in large group settings

Strong Listener-Speaker
- uses story-like sentences describing the beginning, middle, and end of an event
- Uses descriptive words
- consistently contributes in class activities, with relevant responses

READING

Inexperienced Reader
- exhibits limited attention span and little or no experience with books

Early Emergent Reader
- has difficulty grasping print individually
- relies on teacher to read stories aloud
- may still be unaware that meaning comes from text
- participates in shared reading

Emergent Reader
- has growing abilities to use picture clues, memory and patterns to gain meaning from the text

Beginning Reader
- is developing fluency and reads some books with confidence
- able to apply phonics knowledge to decode
- is mostly comfortable reading familiar text with short, simple narrative

Moderately Experienced Reader
- feels comfortable with books
- has confidence reading text independently
- chooses to read for pleasure

Early Childhood Rubric

WRITING

Pre-Writer
- communicates through scribbling and pictures

Emergent Writer
- random letters with drawings
- able to distinguish between drawings and writing
- some evidence of letter-like marks

Developing Writer
- randomly copies words from environment
- demonstrates an understanding that oral language can be written down

Beginning Writer
- beginning to make sound-symbol correspondence
- writing may include familiar words

Experienced Writer
- willing to invent spelling for unknown words
- beginning to write sentences which include key words to communicate thoughts (ct et fd - cat eat food)

Exceptional Writer
- chooses to write independently with an ability to spell some words in a conventional way
- writes complete sentences with punctuation
- shows more consistency in sentence structure, spelling, and vocabulary

APPENDIX B

Language Arts Growth Record

LANGUAGE ARTS GROWTH RECORD

Early Childhood

Name ______________________________ Language Placement ______________

Room ______________________________

ORAL LANGUAGE

Rubrics	Nov.	Mar.	June
Silent/Emergent Listener-Speaker			
Limited Listener-Speaker			
Developing Listener-Speaker			
Capable Listener-Speaker			
Strong Listener-Speaker			

Work Habits	Nov.	Mar.	June
Needs Motivation			
Moderately Motivated			
Motivated			

READING Formal/Informal

Rubrics	Nov.	Mar.	June
Inexperienced Reader			
Early Emergent Reader			
Emergent Reader			
Beginning Reader			
Moderately Experienced Reader			

Work Habits	Nov.	Mar.	June
Needs Motivation			
Moderately Motivated			
Motivated			

WRITTEN LANGUAGE

Rubrics	Nov.	Mar.	June
Pre-Writer			
Emergent Writer			
Developing Writer			
Beginning Writer			
Experienced Writer			
Exceptional Writer			

Work Habits	Nov.	Mar.	June
Needs Motivation			
Moderately Motivated			
Motivated			

Teacher Initial __________ __________ __________

Date __________ __________ __________

For Transfer Purposes: Grade ______

APPENDIX C

Development Rubrics in Mathematics

Ninos Bonitos Elementary School

A Next Century School

DEVELOPMENTAL RUBRICS IN MATHEMATICS

EARLY EMERGENT MATHEMATICIAN

- Communicates/demonstrates little or no understanding of mathematical thinking or mathematical ideas; demonstrates little progress toward accomplishing mathematical tasks.

EMERGENT MATHEMATICIAN

- Communicates/demonstrates partial understanding of mathematical thinking and mathematical ideas. Work may be incomplete or misdirected. Tools and techniques are rarely used.

DEVELOPING MATHEMATICIAN

- Communicates/demonstrates some understanding of mathematical thinking and mathematical ideas. Omissions in conceptual understanding are evident; however, demonstrates the use of mathematical tools and techniques.

MODERATELY EXPERIENCED MATHEMATICIAN

- Communicates/demonstrates an understanding of essential mathematical thinking and mathematical ideas including appropriate representations (i.e.; words, diagrams, graphs, pictures). The work is usually correct and exhibits the appropriate use of mathematical tools and techniques. cooperates with peers and adults.

EXPERIENCED MATHEMATICIAN

- Communicates/demonstrates a sound understanding of essential mathematical thinking and mathematical ideas including appropriate representations (i.e.; words, diagrams. graphs, pictures). The work is usually correct, complete and consistent. May demonstrate the capacity to make generalizations and connections through multiple or unique approaches supported by effective arguments using appropriate mathematical tools and techniques.

EXCEPTIONAL MATHEMATICIAN

- Communicates/demonstrates an in-depth understanding of essential mathematical thinking and mathematical ideas including appropriate representations (e.g.; words, diagrams, graphs, pictures). The work is complete and consistent and demonstrates the capacity to make generalizations and connections through multiple or unique approaches supported by precise logical arguments using appropriate mathematical tools and techniques.

Applications Assessments: Westgate Middle School
June 6-7, 1994
February 16-17, 1995

WESTGATE

APPLICATIONS ASSESSMENTS: WESTGATE MIDDLE SCHOOL

Introduction

The Prince William County (Virginia) Public Schools, as part of an ongoing district-wide restructuring effort, introduced its *Applications Assessments* in mathematics, science, and language arts during the 1993-94 school year. These assessments are just one piece of the school district's restructuring effort, which is guided by a district-developed restructuring process, including a Quality Management Plan, newly adopted Standards of Quality, district-wide school-based management, and other related elements.

The district's work toward a reformed educational system has progressed rapidly. During the summer of 1994, Prince William County teachers worked to develop performance standards for the new Applications Assessments. In addition, teacher teams revising the district's curricula completed their work (in most subject areas), and the new curricula were adopted by the school board. Finally, the district, partially in collaboration with Riverside Publishing Company and partially on its own, developed a fourth Applications Assessment in the area of social studies; this assessment was to be administered for the first time in Spring 1995. Thus, the Prince William County Public Schools provide an example of a district struggling simultaneously with the many pieces of a comprehensive school reform effort.

Westgate Middle School, one of 12 middle schools in the district, was visited twice as a part of this study, first in June 1994 and again in February 1995. The school served 827 sixth, seventh, and eighth grade students during the 1993-94 school year. These students were, on average, somewhat wealthier than the district average, and the school had a smaller percentage of minority students than did the district as a whole. Westgate students were mostly white (80.7 percent); other students were African-American (10.6 percent), Hispanic (4.8 percent), Asian/ Pacific Islander (3.7 percent), and Alaskan/Native American (0.1 percent). About 8.7 percent of students qualify for the U.S. Department of Agriculture's free or reduced price lunch program, a lower percentage than the statewide percentage of students — 12.3 percent — eligible for Chapter 1 services. (These statistics remained essentially unchanged for the 1994-95 school year.) Westgate's faculty includes 65 teachers, two administrators, six teaching assistants, and 25 classified personnel.

Participants

Numerous individuals were interviewed for this case study during the two site visits. Their roles are illustrated in Exhibit 1.

EXHIBIT I

Study Participants

1993-94	1994-95
• Supervisor, Prince William County Public Schools' Office of Assessment and Evaluation • Assistant Principal, Westgate Middle School • Reading Specialist and Assessment Coordinator, Westgate Middle School • Fourteen Teachers, Westgate Middle School — – Eleven seventh grade teachers of math, language arts, social studies, science – One eighth grade teacher of English – Two special education teachers — hearing impaired, emotionally disturbed • Two Special Education Teaching Assistants, Westgate Middle School • Two Parents of Westgate seventh grade students • Sixteen Westgate Middle School seventh grade students — – Thirteen regular education students – Three special education students (hearing impaired), mainstreamed for part of the school day	• Supervisor, Prince William County Public Schools' Office of Assessment and Evaluation • Reading Specialist and Assessment Coordinator, Westgate Middle School • Ten Teachers, Westgate Middle School – Seven seventh grade teachers of math, language arts, social studies, science – Two eighth grade teacher of English – One special education teacher — hearing impaired • One Teacher, member of district's Instructional Support Team • One Parent of Westgate students • Twelve Westgate Middle School eighth grade students

Observations

In February 1995, a team meeting of one of the two Westgate Middle School seventh-grade teaching teams was observed. Additionally, the science classes taught by one of these teachers were also observed.

District Context

Prince William County is a suburban and rural county in Northern Virginia, located not far from Washington, D.C. Many residents commute to the District, but a good portion of the county is rural and far from easy commuting range. The area includes both historical landmarks — several Civil War battles were fought here — and extensive outlet shopping malls.

The Prince William County Public Schools (PWCPS) operate 39 elementary, 12 middle, and 7 high schools throughout the county (the district serves most, but not all, parts of the county; a few small school districts serve the children of some Prince William County municipalities). In all, the district served over 45,000 students during the 1993-94 school year. This student population is approximately 72.0 percent white, 18.8 percent African-American, 5.2 percent Hispanic, 3.6 percent Asian-American, and 0.4 percent from other racial and ethnic backgrounds. Just over 16 percent of the district's students qualify for a free or reduced-price lunch. These statistics remained essentially the same for the 1994-95 school year.

Since the late 1980s, the school district has been planning and implementing a comprehensive restructuring effort encompassing (a) development and adoption of the district's Quality Management Plan, (b) articulation of a set of expectations for student achievement entitled Standards of Quality, (c) expansion of the district's Assessment Program, and (d) revision of the curriculum for all subjects and at all grade levels.

The district's work in these areas is still in midstream, and the restructuring effort, as currently planned, will not be complete for several more years.

Purpose of the Applications Assessments

The Prince William County Public Schools identified a need for a new assessment tool to accompany the district's evolving educational objectives. As part of the district's Quality Management Plan (see below), the district adopted six Standards of Quality, as summarized in Exhibit II. The primary purpose of the Applications Assessment is to help the district measure the progress students make toward the attainment of the new standards.

PWCPS hope to meet at least two additional objectives with the Applications Assessments. First, the assessments will eventually be used for accountability purposes in schools and classrooms, though the district has no plans to introduce rewards or sanctions for performance. However, data will be maintained and reported at the student, classroom, school, and district levels. Second, the Applications Assessments are intended to communicate to teachers, parents, students, and the community what students know and are capable of doing. In this sense, the Applications Assessments join other facets of the district's assessment program to monitor overall student achievement and the quality of the district's educational program.

Although in the future the Applications Assessments will support all of the above purposes, the first full-scale administration in Spring 1994 was intended to collect baseline data about student achievement at grades three, seven, and ten.

Development of the Applications Assessments

PWCPS' Applications Assessments are administered to third, seventh, and tenth graders. The Assessments in mathematics, science, and language arts were piloted in November 1993 and administered district-wide for the first time in May 1994. An Applications Assessment in social studies was developed during the summer and fall of 1994 and was to be administered for the first time in Spring 1995.

EXHIBIT II

Prince William County Public Schools' Standards of Quality[1]

Students should be:

- Knowledgeable and proficient in the traditional basic academic skills
- Good thinkers, problem-solvers, and decision makers
- Effective communicators
- Users of technology
- Knowledgeable of various racial and ethnic cultures, as well as differences based on gender, age, and physical ability
- Good citizens

When the district made the decision to include a performance assessment as part of its Assessment Program, its administrators sought a test publisher that could fulfill the following criteria:

- "Pencil and paper" assessments — administration had to be straight-forward and not overly time-consuming;
- Inclusion in the assessments of higher-order thinking skills and linkages to standards established by such leading organizations as the National Council of Teachers of Mathematics and the Association for Supervision and Curriculum Development;
- Comparability of results to national, normative data; and
- Use of a coherent theme carried throughout the assessment; for instance, on the language arts assessment, the questions to which students respond might all relate to one story, though questions might call for responses employing different types of writing.

Contracting with the Riverside Publishing Company

To develop the first three assessments (in math, science, and language arts), PWCPS contracted out with the Riverside Publishing Company. In the words of the district's Supervisor of the Office of Assessment and Evaluation, "Riverside was the only test publisher that seemed to meet our objectives." PWCPS contracted with Riverside to produce two forms, Form A and Form B, of each assessment. The district will administer the two forms in alternate years (i.e., Form A in 1995, Form B in 1996, Form A in 1997, and so forth).

---

[1] Prince William County Public Schools, *Quality Management Plan*, as revised January 5, 1994.

The Applications Assessments are modeled on the *Riverside Performance Assessment Series* (R-PAS). During the summer of 1993, 180 Prince William County teachers — representing grades three, seven, and ten and language arts, science, and mathematics — met to review and revise potential performance assessment tasks supplied by Riverside. They retained many of the supplied items and made recommendations for revising others. The suggestions made by these teachers led to some changes in assessment items (though not prior to the pilot test of the assessment).

During the summer of 1994, PWCPS began to work with Riverside to select assessments in social studies for grades three, seven, and ten. The district's administrators and teachers felt that the assessments Riverside offered for grades three and seven were appropriate for the district's students and curriculum but that no assessments available through Riverside were appropriate for the district's tenth graders. Consequently, the district opted to develop its own Applications Assessment for tenth grade social studies. During the 1994-95 school year, teacher volunteers worked to develop the assessment. Riverside is assisting the district in its work by critiquing the assessment; ultimately, Riverside will help the district revise and package the assessment.

Pilot Tests

The Form A Applications Assessments in math, science, and language arts were piloted in November of 1993. All high schools and middle schools and about half of the district's elementary schools participated in the pilot test; approximately 400 students at each grade level were included in the pilot. (Students participating in the pilot were fourth, eighth, and eleventh graders, as these students had only recently completed the grades at which the district administers the assessments.) The pilot test was scored by two corporations hired by Riverside Publishing Company.

During the winter of 1994, the district reviewed the results of the pilot test with teachers. Additional assessment items also were evaluated, and, again, the district provided Riverside with recommendations for modifications prior to the spring administration of the test.

In November 1994, Form B Applications Assessments in math, science, and language arts were piloted. In addition, the new Applications Assessments in social studies were piloted for grades three and seven. The tenth-grade social studies Applications Assessment was to be piloted later in 1995.

Characteristics of the Applications Assessments

The Applications Assessments constitute a performance event during which students must construct responses to questions and prompts in the areas of language arts, mathematics, and science. During the Spring 1994 administration of the exam, each student took only one of the three sections. For the Spring 1995 administration, all students were going to be required to take the math and the language arts assessments. In addition, the district was going to administer the science and social studies assessments to a sample of 25 percent of students (25 percent of students were to take the science assessment, 25 percent the social studies assessment, and the remaining 50 percent would take neither).

Format and Time Limits

Each assessment is contained within a booklet. A typical assessment has about a dozen items which call for anything from a phrase to a multiparagraph or multistep response. In 1994 students were allowed up to two hours to complete the exam, although most students reportedly took much less time to finish, and administrators have called the exam untimed. For the Spring 1995 administration, the assessment will be essentially untimed; that is, students actively working to complete the assessment may have as much time as they require. The assessments, however, will still be designed to take most students about two hours to complete.

Scoring Rubrics

Scoring rubrics are used on each item on the Assessments and are tailored to the individual task; the rubric for one task might call for the scorer to rate students' responses on a 1- to 4-point scale, while other items might be worth only 1 point — the student either gets credit for the response or does not. Most rubrics for individual assessment items differentiate between responses on 2-, 3-, or 4-point scales. (Examples of two rubrics — one for a series of math tasks and one for a series of language arts tasks — appear in Appendix A.) Scoring of the Applications Assessments is done by an outside company associated with Riverside. At one point in time, the district considered having its own teachers score the assessments, but gave up the idea as impractical.

Provisions for Special Populations

Students who receive special education services take the Applications Assessments under modified conditions, as called for by each student's IEP (or not at all, again, according to the IEP). Mediated conditions include extra time, extra instructions from the teacher, dictating responses to an adult to write down, and assistance in reading the assessment items.

Development of Standards of Performance for the Applications Assessments

Over the summer of 1994, Prince William County teachers came together develop standards of performance for the assessments. The approach taken by the district to accomplish this attack was two-pronged. First, without having any information available to them concerning actual student performance on the spring 1994 assessment, teachers wrote standards for satisfactory performance on the assessments at each grade level. Then, after reviewing information about the scored assessments, teachers reviewed the standards they had constructed to reconsider whether or not they were appropriate.

Based upon the standards written by Prince William County teachers, between 40 and 60 percent of students assessed in Spring 1994 met or surpassed the standards for the nine assessments (i.e., third-grade math, science, and language arts; seventh-grade math, science, and language arts; and tenth-grade math, science, and language arts). According to the Supervisor of Assessment and Evaluation, teachers adhered to the standards they had written even in the cases where fewer than half of students assessed met the standard.

The district's Supervisor of Assessment and Evaluation was, he said, very pleasantly surprised by how well the standard-setting process went. The process went more smoothly than he would have anticipated, and the standards written were of high quality. He also reported that teachers were equally pleased with the process and its outcomes.

Evaluations of the Applications Assessments

Riverside Publishing Company conducted studies of the interrater reliability of the scoring process following both the 1993 pilot administration and the Spring 1994 full-scale administration of the Applications Assessments. In both cases, they found interrater reliability to be high; for the 10 percent sample of Spring 1994 Applications Assessments for which interrater reliability was computed, Riverside found it to be at about 95 percent.

District administrators assert that the assessments were chosen by virtue of their content validity: the district believes they assess something valuable. However, the district intends to continue to review the content validity of the assessments with teachers.

Teachers have expressed two types of concerns about the content validity of the assessments. First, math teachers, both at the tenth-grade and the seventh-grade levels, are concerned that the assessments include some material many students have not had a chance to study or to master (for instance, the tenth grade assessment includes some geometry, and the seventh grade assessment focuses on percentages). Second, the content validity of the Applications Assessments is necessarily ambiguous given the fact that the district selected and implemented the assessments prior to revising the curriculum. Thus, the relationship between the Applications Assessments and the revised curriculum remains to be examined.

District administrators say they intend to bring teachers together to review these unresolved issues. In addition, the district intends to follow up with teachers concerning the consequential validity of the assessments — that is, how teachers use the information provided by the Applications Assessments. These evaluations, however, remain to be done.

Resource and Staff Development Support

To develop the Applications Assessments, PWCPS has drawn on financial and human resources from both within and outside the district. In addition to supporting the development of the assessments, these resources have supported several staff development activities designed to inform teachers about the new assessments and to solicit their reactions. Says one teacher, "The district office person coming to school and giving us an inservice was one of the things done right." In addition, the district has attempted to keep the community informed about the assessments and their purpose.

Financial Resources

The development of the Applications Assessments will have cost the PWCPS approximately $300,000 over the 1993-94 and 1994-95 school years. In addition, expenses associated with scoring the assessments cost $4.50 per student; as approximately 10,000 students in grades three, seven, and ten take the assessments each year, annual scoring costs will be about $45,000.

Approximately 25 percent of the cost of the program is borne by the Federal government, while the remaining 75 percent comes from local sources.

Human Resources

Many PWCPS teachers, especially those who teach third, seventh, or tenth grade, have been involved in the process of refining the Applications Assessments. Additionally, many high school social studies teachers have taken on the job of developing the tenth-grade social studies Applications Assessment. The involvement of the district's teachers represents both a source of human resources as well as staff development, as teachers learn more about performance assessments and return to their schools with their new knowledge.

During the winter of 1994, the district held two two-day workshops for (primarily) third, seventh, and tenth grade teachers. About 70 teachers attended each workshop, and 20 more attended a make-up session. The workshops focused on the following:

- An overview of the Prince William County Assessment Program and the role performance assessment will play in it;
- Teachers taking the assessments to be administered to students at the grade level at which they taught;
- Teachers learning how to use the scoring rubrics and scoring their own exams. As part of reviewing the rubrics, teachers also saw anchor papers to help identify low and high quality responses;
- A presentation of results of the November pilot test, item by item;
- Discussions of how test items could be improved in the future;
- Discussion of how the next step in the restructuring effort — the standard setting process — can best be approached;
- Planning for the spring administration of the Applications Assessments; and
- Review of the Resource Kit (see below) to be distributed to all schools.

Teachers who attended these workshops were charged with the task of bringing what they had learned back to other teachers at their schools.

The district did not anticipate repeating these workshops prior to the Spring 1995 administration of the Applications Assessments.

Teacher Resource Kit

The district compiled a Resource Kit, which was distributed to all schools in the district in March of 1994. The Resource Kit provides teachers with information about both the Applications

Assessments and the Basic Skills Assessments, the two new elements in the district's Assessment Program. A similar Resource Kit was to be distributed to all schools prior to the Spring 1995 administration of the assessments.

School Focus and Videos

Two district administrators, the Supervisor of the Office of Assessment and Evaluation and the Assessment Specialist, appeared on a local public television program called *School Focus*, in January of 1994, to discuss the new Applications Assessments and other elements of the district's assessment program. The district recorded this program on video for inclusion in the Teacher Resource Kit and for broader distribution to parents and community members. Similarly, district administrators made another video, this one concerning the district's entire assessment program — norm-referenced tests (ITBS), criterion-referenced tests, Applications Assessments, and Basic Skills Assessments — in February 1995. This video, too, was to be made available to district teachers and other interested individuals.

Interaction with Other Reforms

The Applications Assessments are only one of several reforms Prince William County is currently introducing into its school system. The district has adopted a Quality Management Plan, which delineates a plan for restructuring schools to support students' attainment of six Standards of Quality adopted by the district (described above). One of the six standards focuses on basic skills, and the district is adopting new Basic Skills Assessments to complement the Applications Assessments. In addition, the district is in the process of revising its curricula at all grade levels, in all subject areas. Finally, a state-level reform — Virginia's Literacy Passport — is a significant piece of the district's Assessment Program.[2]

Quality Management Plan

The Prince William County Public Schools Quality Management Plan (QMP) was developed over two years, beginning in the 1991-92 school year. Borrowing the language of its exemplar, Total Quality Management (the plan speaks in terms of "customer satisfaction"), the plan sets out a vision statement, a mission statement, Standards of Quality, goals, and performance standards for the school district in order ". . . to define a single purpose for all employees, to focus on that purpose, and to continuously improve the operation of Prince William County Public Schools" (QMP, p. 1). The QMP encompasses the school district's long-range plan, schools' annual plans, school district department plans, and the district budget. Adopted in June of 1993, the QMP was revised in September of 1993 and again in January of 1994.

---

[2] One additional element of restructuring initiated at the district level is school-based management. This piece of the district's restructuring effort is described below because, although it was initiated at the district level, it is carried out primarily at the school level.

Standards of Quality

One important element of the district's Quality Management Plan is the six Standards of Quality, already described above. During earlier iterations of the QMP, only five standards were delineated. However, concern in the community over what some people perceived to be value-laden standards led, in January of 1994, to the addition of a standard for students' basic skills attainment. The Standards of Quality are what the district's Supervisor of Assessment and Evaluation calls the district's "expectations" for student educational outcomes, and it is progress toward the attainment of these standards that the Applications Assessments and the Basic Skills Assessments are intended to measure.

Basic Skills Assessments

The second recent addition to Prince William County's Assessment Program is the Basic Skills Assessments. Piloted in the Spring of 1994 in conjunction with the first full-scale administration of the Applications Assessments, the Basic Skills Assessments are multiple-choice, criterion-referenced exams administered at grades 3, 5, 7, and 10. These assessments are intended specifically to measure student progress toward attainment of the Standard of Quality pertaining to "basic skills."

Applications Assessments and Basic Skills Assessments are intended to complement each other and other components of the district's assessment program (e.g., the ITBS and Virginia's Literacy Passport). While the Applications Assessments aim to measure students' abilities to apply knowledge, the criterion-referenced Basic Skills Assessments are more traditional in their assessment of students' mastery of subject matter.

Curriculum Revision

During the 1993-94 school year, the Prince William County Public Schools moved to revise its curriculum at all grade levels. Teachers interested in participating met to evaluate the existing curriculum and to identify desirable revisions. Groups of teachers (e.g., fourth grade teachers, seventh grade science teachers) moved at different speeds, but over the course of the school year, most groups presented preliminary revisions of the curriculum to the group(s) they were representing. Teachers district-wide were then invited to comment on the revisions. Over the Summer of 1994, the revisions were essentially complete, and the school board formally adopted most of the revised curricula.

The implication of curriculum revision for the Applications Assessments is unclear. One teacher interviewed for this study has been involved in the revision of the language arts curriculum; she suggests that the revised curriculum corresponds well to the Applications Assessment in language arts. However, the Applications Assessments, according to administrators and other teachers, did not guide the curriculum revisions in any direct way. Therefore, it remains for the district's future evaluation of the content validity of the assessments (described above) to reveal how closely the revised curricula and the Applications Assessments correspond to one another.

Virginia's Literacy Passport

The state of Virginia introduced its *Literacy Passport* in 1989. The assessment, taken by all sixth graders, comprises three sections — reading, writing, and mathematics.

Successful completion of all three sections is intended to reflect students' mastery of basic literacy. Students who do not pass the exam during either of the two sixth grade administrations may retake those sections they do not pass during their seventh and eighth grade years. Students who still have not passed all three sections by the end of eighth grade are prohibited from being considered "full" ninth graders; although they may be promoted to high school, they may not take part in certain extracurricular activities and elective coursework until they have passed all three sections of the exam.

Because of the fairly high-stakes nature of the exam, the Literacy Passport has been somewhat controversial. The high stakes also mean that it is a well-known rite of passage at Westgate Middle School. Teachers, students, and parents alike are familiar with the exam and its consequences.

School Context

Westgate Middle School, located in suburban Prince William County, is housed in a building formerly used as a high school. The school is marked by a decentralization of authority — the principal is a firm believer in delegation of responsibility. The school, like other Prince William County schools, has adopted site-based management practices.

Since the late 1980s, seventh and eighth grade teachers have been organized into teams, each grade level being divided into two groups. Thus, one half of the seventh grade student body is taught by teachers on Team A, while the other half is taught by Team B teachers. Teaching teams have common planning time every other day, one group meeting on even numbered days and the other on odd.

School Level Education Reforms

In addition to the elements of education reform introduced by the district and described above, Westgate Middle School itself has initiated or been a central player in two school-level reforms. The first is school-based management, initiated by the district over the past five years but carried out at the school level. Second, Westgate Middle School has been introducing portfolio assessment school-wide during the past three years.[3]

---

[3] Woodbridge Middle School's use of portfolios, though an element of the restructuring taking place at the school, is not the focus of this case study. Thus, they are described only briefly here.

School-based Management

Beginning in 1989, the Prince William County School District piloted and then introduced districtwide school-based management. Westgate Middle School is managed by a Planning Council comprised of 14 teaching and other instructional staff members, 6 parents, and 2 administrators.

All schools in Prince William County must adopt an Annual School Plan. Westgate's Annual Plan for the 1994-95 school year (adopted in January 1994) sets out eight objectives to be achieved during the school year. One of the eight objectives is "To demonstrate growth and improvement based on academic achievement, surveys, reports, portfolios and standardized test results as measured by Iowa Tests of Basic Skills and Virginia Passport Testing" (Westgate Middle School, Annual School Plan, January 31, 1994, p. 21).

Interestingly, though perhaps not surprisingly, conflicting impressions of the balance of power on the Planning Council emerged from individuals participating in this study. While one teacher was indignant that, "We allow parents to control this facility. . . . They dominate the school council," a parent suggested that parents have very little influence at all.

Portfolios

The aforementioned reforms are all district driven reforms. At the school level, however, another assessment reform is taking place. Throughout the school, teachers are helping their students to maintain portfolios of their work over the course of each school year.

During the 1991-92 school year, interested teachers at the school, led by the Language Arts Department Head, researched the portfolio concept and worked on developing a method of using them that made sense to Westgate teachers. During 1992-93 seventh and eighth grade teachers used portfolios for the first time. Each team develops a portfolio system that works for them. Every student has a faculty advisor from within his or her team who is responsible for guiding the student in the selection of work to include in the portfolio; each teacher advises about 17 students. Quarterly, students select work they have completed for inclusion in the portfolio, conduct self-assessments of their work, and develop goals for the upcoming quarter.

Westgate teachers do not score the portfolios. They do suggest, however, that the portfolios have provided valuable insights into their students' learning processes and attitudes toward school. For instance, one teacher commented, "When the kids chose almost only tests to go into their portfolios the first quarter, it really shook us up. We realized we had to look for ways to encourage students to value other types of work." Additionally, teachers suggest that they have a broader understanding of their students than they used to have when they never saw the students' work in other subject areas.

Impact of the Applications Assessments and Other Reforms

Westgate Middle School provides an example of a school in transition. Multiple reforms are taking place simultaneously, and their ultimate impact — singly and cumulatively — remains to be seen. It is too early to discern an impact of even single elements of districtwide and schoolwide reform, such as the Applications Assessments. Still, the reactions of teachers, students, and parents to the Applications Assessments and other reforms can be described.

Impact on Teachers

Teachers at Westgate Middle School fall into two camps over the Applications Assessments and other current restructuring efforts — the generally optimistic and the somewhat skeptical.

Teachers' reactions to the applications assessments. Most Westgate teachers hold favorable opinions of performance assessments in theory. As two teachers commented, "It's the only logical way of doing assessment," and "We need to get these kids to write more. They've been allowed to fill in dots for too long."

Many teachers' reactions to the Applications Assessments are also positive, though some hold more critical views. Teachers found that the language arts and science assessments were well aligned with the district's current curriculum and actual classroom experiences, and one language arts teacher felt the assessment was also well aligned to the new curriculum. However, several math teachers said that the math assessment was developmentally inappropriate for their students, for it tested skills most seventh grade students have not yet mastered (much of the exam focused upon the calculation of percentages). Teachers also expressed concern about inaccuracies in the math assessment; the assessment misused the words "percentage" and "percent," for example, thus confusing students who were aware of the distinction.

Some teachers expressed concerns that the science and math assessments were really reading comprehension tests, noting that "kids who don't read well have a problem." One math teacher felt that the math Applications Assessment represented an encroachment of language arts into the mathematics discipline; this teacher said, "If it's something that helps students learn math, I'll do it," but that she was not yet certain that "all this writing in math" is beneficial to student learning. Another math teacher, however, expressed very positive feelings about the math assessment, noting that it is aligned with the standards developed by the National Council of Teachers of Mathematics.

Finally, one teacher expressed a concern about the subjectivity inherent in the use of scoring rubrics, and another cautioned, "Performance assessment is not the end all and be all. Educators tend to get caught up in fads."

Teachers' reactions to the development process and training activities. Teachers applauded the district's efforts to include them in the process of developing the Applications Assessments. However, one teacher who had been involved in the refinement of test items indicated that Riverside had not incorporated the changes she and other teachers had recommended for one of the assessments. Another teacher expressed dismay that the district wanted her to serve as a "rubber stamp;" in her opinion "the district had already chosen a testing company to create a test for a curriculum we hadn't chosen yet."

Teachers who had participated in some form of inservice activity focusing on the Applications Assessments were universal in their praise of the format and usefulness of these activities. Special education teachers and aides also reacted favorably to the inservice activity that focused on how they could provide accommodations for their students taking the assessments.

Teachers' reactions to other reforms. Most Westgate teachers expressed positive reactions to the curriculum revisions they had seen as of spring 1994. One language arts teacher liked the fact that the new curriculum placed "more emphasis on *how* than on *what*" teachers taught and students learned. Math and science teachers also expressed satisfaction that their curricular objectives were being reduced to a much more manageable number. One teacher, however, said, "Restructuring is not the answer. It's cyclical. We've been through it all before. These curricular objectives look a lot like the ones we had 12 years ago."

Westgate teachers expressed one substantial dissatisfaction with the process the district is pursuing to implement the new standards and other reforms. Because of the reform effort, the district has reassigned its subject area supervisors to Restructuring Support Teams (renamed Instructional Support Teams in 1994-95). A direct byproduct of the expanded Assessment Program and other reforms, this reassignment of district personnel has left some teachers feeling rudderless. As one teacher put it, "Introducing a revised curriculum and eliminating our area supervisors at the same time is a double whammy." This widespread reaction on the part of Westgate Middle School teachers suggests a point of disconnect between district officials and teachers. Though district administrators maintain that the supervisors are still available and merely are wearing different hats, this suggestion is clearly at odds with teachers' perceptions, expressed both in 1994 and in 1995.

Teachers' comments on feedback and communication with district. Surprisingly, Westgate teachers reported in February 1995 that they had not yet received feedback from the district on student performance on the Spring 1994 Applications Assessments. Additionally, most teachers said they were unfamiliar with the standards of performance a group of district teachers had developed over the summer.

Teachers were also ignorant about other facts concerning the Spring 1995 administration of the assessments. For instance, teachers interviewed did not know that students would take both the language arts and the math assessments in 1995 and that a subset would also take the science or social studies assessment (in 1994 each student took only one of the assessments). Furthermore, teachers were unaware that the math, science, and language arts assessments to be administered in 1995 would be identical to those administered in 1994 (Form A was used in 1994 and 1995; Form B will be used in Spring 1996 for the first time). (Note, however, that at the time of the 1995 interviews, the district had not yet distributed Resource Kits for the 1995 administration of the Applications Assessments.)

Impact on Curriculum and Instruction

A few teachers identified small changes in their instructional practices made in response to the introduction of the Applications Assessments. For example, one math teacher said that, because of the emphasis of the 1994 assessment on percentages, she had rearranged her curriculum so as to cover percentages earlier in the school year (she reported that last year's assessment was administered before she had taught percentage concepts to students). A special education teacher also said that, because of her experience administering the Applications

Assessments to her hearing-impaired students, she had begun to reflect upon changes she could make in her instructional methods to foster better language development in these children. However, she said she had not yet made any changes in her instructional techniques.

Most Westgate teachers identified only modest changes in their approaches to instruction and assessment in their classrooms. An illustration of one teacher's classroom techniques and changes compatible with, though not necessarily spurred by, the Applications Assessments appears in Exhibit III.

However, the predominant impact the assessments have had on instruction to date is in the area of test preparation. Teachers prepared their students for the assessments in different ways. Some spent as much as three days prepping students, while others spent only a few minutes. Thus, the number of school hours devoted to test preparation and administration varied across classrooms. Several teachers purchased sample tests, paying for them out of their own pockets, to use with their students.

Two teachers stated that the Applications Assessments will necessarily affect instruction. While one teacher linked the impact of the Applications Assessments on instruction to any future accountability system ("When we have accountability, we will teach directly to the test."), another prophesied, "If nothing else, there will be another kind of test-taking skill to be taught."

Impact on Students

In 1993-94, students' almost universal responses when asked about the Applications Assessments were that they were "easy" and that they prefer multiple-choice exams. Some students said they liked having the chance to explain their opinions about something (on the language arts exam), while others said that the themes carried throughout the exams were "boring." One girl said, "They should try to choose a theme that kids will like better," reacting to the math assessment's "crossword puzzle contest" theme.

Despite students' assertions that the assessments were easy for them, their teachers doubt that they performed well. According to one teacher, "We're going to have to lower our standards and acknowledge that these children have not been trained to be problem solvers. . . . Ninety percent of our kids can't do this kind of test. It's an idealistic assessment." Teachers suggested that their students did not take the test seriously; they saw no reason to put in their best effort because they were not being graded on it and no score would go home.

In 1994-95, seventh-grade students had not yet taken the Applications Assessments, and the eighth-grade students interviewed expressed similar reactions to those shared by their classmates the previous year.

Impact on special populations. Students receiving special education services had an especially hard time with the Applications Assessments. As one Teaching Assistant put it, "The Applications Assessments focus on critical thinking skills and problem solving skills — things these kids usually struggle with. They also have trouble with reading comprehension, which is also a problem on this kind of test."

EXHIBIT III

One Teacher's Reflections Upon Education Reform and Modifications in Her Classroom

Suzanne Conway runs a busy, fast-moving seventh-grade life science classroom. During five 50-minute class periods each day, she daily makes minor changes in the day's lesson plan so that she can keep things interesting and test out teaching techniques that work best.

Though Suzanne says she has not made any changes in her teaching practices as a response to the district's introduction of the Applications Assessments, her teaching has changed over the past several years in ways compatible with the new assessments. For instance, Westgate Middle School's adoption of team teaching several years ago has led to increased (though, Suzanne says, by no means complete) integration of subject area instruction. She says she emphasizes written language much more in her class than she did in the past, and she also introduces students to multiple ways of communicating information, such as models, graphs, charts.

Suzanne says she has used rubrics for at least 15 years to score students' project work. However, students do not see the rubrics ahead of time, and Suzanne believes that she could improve her skills in writing and using rubrics.

Most directly relevant to the district's new assessment is Suzanne's recent emphasis on what she calls "Science Starters." These activities, described on the blackboard of her classroom, emphasize science skills (e.g., data collection techniques) and concepts (e.g., cellular reproduction), not rote memorization of information.

Students enrolled in a class for children with hearing impairments commented that they prefer multiple choice tests to the Applications Assessments (most of their nondisabled peers said the same thing). However, those same students liked the fact that the new assessment allowed them to ask their teacher to provide them with extra directions. That was "a good idea," said one boy. The teacher of the children with hearing impairments commented that the Applications Assessments are not a valid test of their knowledge because the emphasis on language makes it so hard for them; the students must read, and then write, making the Assessments doubly hard for them. This teacher speculated that she would probably place more emphasis on the types of questions she uses with her students, aligning them with the Applications Assessments.

Impact of portfolios. In both 1993-94 and 1994-95, students' responses to the use of portfolios at WMS are almost universally positive. Though a few students grumbled about having to go over work they had already completed, most agreed with one boy who said, "I like looking back at my past work and seeing how much I've improved."

Impact on Parents

The impact of the Applications Assessments on parental involvement in children's education is, as of yet, negligible. The two parents who participated in this study were not well-informed about the Applications Assessments. They were aware that their children had taken the exam, and they knew which of the three sections they had taken. Interestingly, neither parent believed she had received any information from the school about the assessment ahead of time, despite the fact that a notice was sent home. (One parent chalked this up to the unreliability of her seventh grader cum messenger.)

One of the parents said that her daughter has significant trouble taking tests. The seventh grader has not yet passed all three sections of the Literacy Passport, and both mother and daughter are concerned about her being able to participate as a "full" ninth grader when she enters high school.

Future Plans

The Prince William County Public Schools intend to use the results from the spring 1994 administration of the Applications Assessments as baseline data against which future student scores can be judged. The district has waffled in the past about the level(s) at which scores would be reported (at one point, the district anticipated reporting only district-wide scores), but as of February 1995 administrators seemed committed to reporting student, classroom, school, and district level scores, enabling schools and teachers to monitor how they are doing compared with their counterparts throughout the district.

As has been described above, the district anticipates conducting ongoing evaluations to the Applications Assessments' content and consequential validity, and district administrators envision teachers playing important roles in these evaluations.

Conclusions

The Prince William County Public Schools has embarked upon a comprehensive restructuring effort that will require several more years to bring to fruition. As has been noted above, the singular and cumulative effects of various reforms — including the Applications Assessments — cannot yet be determined.

In embarking upon a comprehensive education reform effort — encompassing extensive changes to curriculum, assessment, and management districtwide — the Prince William County Public Schools have clearly set themselves an ambitious task. The rapid introduction of multiple elements of reform has occasionally resulted in some difficulty in coordinating the various pieces: changes in any one element of education reform tend to have ramifications for other elements, and the greater the number of reforms going on simultaneously, the greater the number of effects.

This is not to say that the Prince William County Public Schools has failed to plan its education reform adequately. To the contrary, the Quality Management Plan is just that — a plan — and any reform effort of substance must be flexible if it is to survive. So far, the district has been able to continue on the road to restructuring, even while the map it is following continues to be drawn.

APPENDIX A

Sample Scoring Rubrics

Riverside Performance Assessment Series
Scoring Guide for Reading: The Day of the Fire
Level D

This page may be photocopied and used to record and report student scores on the assessment.

The conventions of writing, such as capitalization, punctuation, and usage, should not be a distraction in responses at score points 3 and 4.

Assign a 0 if the student does not attempt the exercise or fails to meet the requirements set for score point 1. Assign an N/S (Not Scorable) if the student's work is indecipherable or written in a language other than English.

Student Name: ______________________ Grade: __________

School: ______________ Teacher: ______________________

No.	Scoring Guidelines and Acceptable Responses	Maximum	Score
1	**Award 1 pt. for each correct answer.** choice 3: "We wanted to build . . ."	1	
2	**Accept similar wording.** The fire fighters burned down the old home.	1	
3	This exercise should be scored in two parts. For the line or bar section, assign 1 point if the bar is marked near the middle. (Imagine the bar divided into thirds and give credit if the mark appears within the middle third.) For the second part of the exercise, assign 1 point if the student refers in some way to the ambivalent emotions the boy reveals through his actions, through what he says, through his reactions to his surroundings, or through his direct description or explanation. It is possible for a student to receive credit for one section of the exercise and not the other.	2	
4	Assign 1 point if the student correctly identifies the probable event, "The author's family builds a new house." Assign 1 point if the student supports his or her choice with an appropriate reason, such as "The boy says they wanted to build a new house, but they had to get rid of their old house first."	2	
5	Assign 1 point for each response that provides reasonable support for the student's choice, even if the emotion is not clearly implied by the narrative. For example, "embarrassed," although not implied in the narrative, is not unreasonable if supported by the idea that the boy may have been embarrassed by all the curious attention given by neighbors to the event. Even one of the least supportable responses, such as "jealous," should be given credit if a rational reason is given for this choice (e.g., "He probably felt jealous because other people can move out of a home and still be able to go back and see it again").	2	
6	Assign 1 point if the student correctly completes the summary, using words similar to the following: A boy feels mixed emotions (happy and sad) as he watches his old home being burned down to make way for a new one.	1	

BEST COPY AVAILABLE

© 1993 The Riverside Publishing Company. This page may be photocopied.

No.	Scoring Guidelines and Acceptable Responses	Maximum	Score
7	A 3 response clearly identifies a similarity or difference between the author and the student and explains how this relates to the boy's actions and feelings in the narrative. Even a seemingly insignificant difference, such as "He is a boy and I am a girl," can be part of a 3 response if it is related to different attitudes or behavior. A 2 response accurately identifies a similarity or difference, but fails to clearly relate this difference to the boy as he appears in the narrative. A 1 response attempts to answer the question, but fails to clearly identify a significant difference or similarity.	3	

© 1993 The Riverside Publishing Company. This page may be photocopied.

Riverside Performance Assessment Series
Scoring Guide for Mathematics: Day Camp Dimensions
Level E

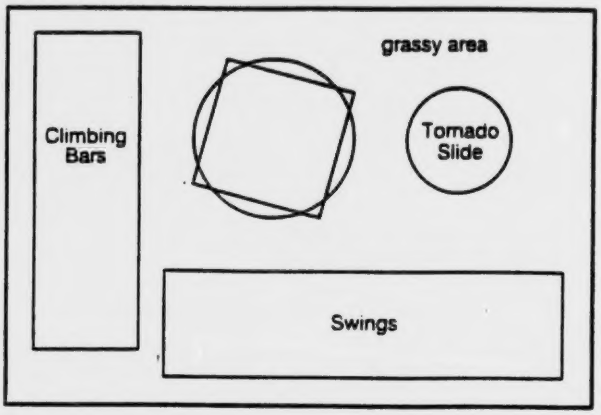
This page may be photocopied and used to record and report student scores on the assessment.

The conventions of writing, such as capitalization, punctuation, and usage, should not be a distraction in responses at score points 3 and 4.

Assign a 0 if the student does not attempt the exercise or fails to meet the requirements set for score point 1. Assign an N/S (Not Scorable) if the student's work is indecipherable or written in a language other than English.

Student Name: ____________________ Grade: ____________

School: ____________________ Teacher: ____________________

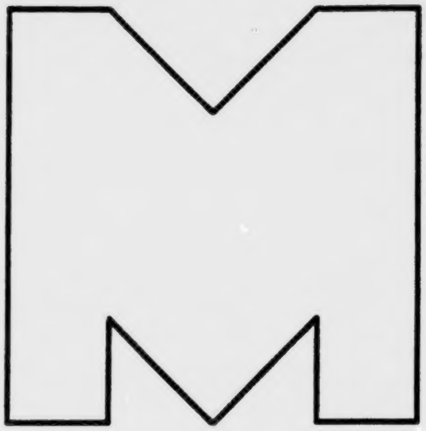
No.	Scoring Guidelines and Acceptable Responses	Maximum	Score
1	Note: Adequate responses meet all of the following criteria: (a) A square is shown. (b) A circle is shown. (c) The sides of the square are not parallel to any side of the playground. (d) The circle is no closer than 10 ft to any playground equipment or the edge of the playground. (e) The square is no closer than 10 ft to any playground equipment or the edge of the playground. A typical drawing is shown below.  A 3 response meets all five criteria given above. A 2 response meets three or four criteria given above. A 1 response meets only one or two criteria given above.	3	

© 1993 The Riverside Publishing Company. This page may be photocopied.

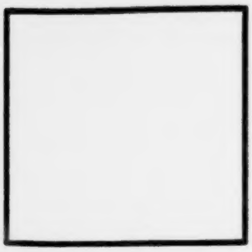
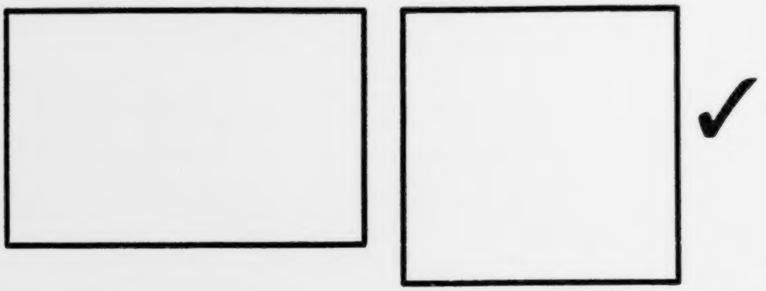
No.	Scoring Guidelines and Acceptable Responses	Maximum	Score
2	Note: Allow for faulty input data from number 1. A student may receive full credit on this item for properly estimating the areas of a circle and a square that did not meet all of the specifications for number 1. If one of the two required figures is missing from number 1, the highest number of points that can be earned for this number is 2. If the student failed to respond to number 1, score a **0** for this item. Acceptable responses meet all of the following criteria: (a) The formula $A = \pi r^2$ is used to calculate the area of the circular sandbox. (b) A reasonable estimate of the area of the square sandbox is given. The estimate may have been arrived at through use of the formula $A = bh$ or through counting squares. (c) All measures used in the formulas match the measures shown in the diagram. (d) No computational errors are shown. A 4 response meets all four criteria given above. The response may show the correct units. A 3 response meets criteria (a), (b), and (c) but fails to meet criterion (d). A 2 response meets criteria (a) and (b) but fails to meet criterion (c) and perhaps (d). A response that fails to meet criterion (c) may use diameter in the formula instead of radius or overlook the fact that each square on the grid represents 25 square feet. Alternately, a 2 response meets either criterion (a) or (b) (not both) and criteria (c) and (d). A 1 response meets either criterion (a) or (b) (not both) and may fail to meet criteria (c) and (d).	4	
3	Note: Allow for faulty input data from number 2. Students can receive full credit for this item using areas calculated incorrectly in number 2. A 3 response selects the sandbox with the greater area (as calculated in number 2), determines the volume of this sandbox by multiplying the area by 2 ft, compares the resulting volume to the 1400 ft^3 minimum for delivery by truck, and selects the best way to order the sand. A 2 response correctly determines the volume of one of the sandboxes using input data from number 2, but it shows at least one of the following errors: calculations show computational errors; the sandbox with the lesser area is selected; the wrong method of ordering sand is selected. A 1 response employs a flawed method for determining the volume of one of the sandboxes and selects a method for ordering sand based on the calculations.	3	
4	A 3 response explains a way to make a compass and use it to create a circle with a radius of 6 ft. The explanation is clear, lists all of the materials that would be needed, and shows no gaps in logic. A 2 response explains a way to make a compass and use it to create a circle with a radius of 6 ft, but the explanation shows minor flaws. For example, the response may not explain how to secure the center point or how to mark the circle on the grass. A 1 response explains a method for making a circle on the grass, but the method is unclear and/or shows serious flaws. Alternately, a 1 response explains a way to make a compass but either neglects to mention the radius of the circle or suggests a radius of 12 ft.	3	

BEST COPY AVAILABLE

© 1993 The Riverside Publishing Company. This page may be photocopied.

No.	Scoring Guidelines and Acceptable Responses	Maximum	Score
5	Note: A sample design is shown on the grid below.  Acceptable responses meet all of the following criteria: (a) The letter M is diagrammed on the grid. (b) The height and width of the M equal 12 ft. (c) The design reflects an effort to maximize the space available for planting vegetables within the M. A 3 response meets all three criteria given above. A 2 response meets only two criteria given above. A 1 response meets only one criterion given above.	3	
6	Note: Allow for faulty input data from number 5; that is, give students credit for computing the perimeter of a garden from number 5 that did not meet all of the specifications. Acceptable responses meet all of the following criteria: (a) The formula $C = \pi d$ (or $C = 2\pi r$), where $d = 12$ ft, is used to calculate the circumference of the circular garden. (b) A reasonable estimate of the perimeter of the M-shaped garden is given. (c) No computational errors are shown. Typical calculations for the circumference of the circle are shown below. Acceptable responses need not be rounded as shown. $C = \pi d$ $= 3.14 \times 12$ ft $= 38$ ft Estimates for the perimeter of the M-shaped garden that are between 55 and 60 feet are acceptable. A 3 response meets all three criteria given above. The response may show the correct units. A 2 response meets criteria (a) and (b) but fails to meet criterion (c). A 1 response meets either criterion (a) or (b) (not both) and may fail to meet criterion (c).	3	

© 1993 The Riverside Publishing Company. This page may be photocopied.

<table>
<tr><th>No.</th><th>Scoring Guidelines and Acceptable Responses</th><th>Maximum</th><th>Score</th></tr>
<tr><td>7</td><td>Note: The correct response is shown on the grid below.

A 2 response shows a square tray measuring 6 tiles by 6 tiles.
A 1 response employs all 36 tiles but does not show a square tray. For example, the next best arrangement would be 9 units by 4 units.</td><td>2</td><td></td></tr>
<tr><td>8</td><td>Note: A typical response is shown on the grid below.

A 3 response shows two trays with any of the dimensions shown in the table below. The response shows a check by the tray that requires the lesser amount of border, unless both require the same amount, in which case both or neither may be checked.
<table>
<tr><th>Dimensions</th><th>Number of tiles</th><th>Border length</th></tr>
<tr><td>6 × 9</td><td>54</td><td>30</td></tr>
<tr><td>5 × 10</td><td>50</td><td>30</td></tr>
<tr><td>7 × 7</td><td>49</td><td>28</td></tr>
<tr><td>6 × 8</td><td>48</td><td>28</td></tr>
</table>
A 2 response includes one or two trays with dimensions other than those given above, although the area of each is greater than 41 units^2 and less than 55 units^2. The response shows a check by the tray that requires the lesser amount of border, unless both require the same amount, in which case both or neither may be checked. Alternately, a 2 response shows two trays with dimensions from the table above but marks the tray requiring the greater amount of border.
A 1 response includes one or two trays with dimensions other than those given above, although the area of each is greater than 35 units^2 and less than 61 units^2. The response also marks the tray requiring the greater amount of border.</td><td>3</td><td></td></tr>
</table>

© 1993 The Riverside Publishing Company. This page may be photocopied.

Vermont's Portfolios:
Maple Leaf Middle School
May 10-11, 1994
March 16-17, 1995

VERMONT'S PORTFOLIOS:
MAPLE LEAF MIDDLE SCHOOL

Introduction

The focus of this case study is the Vermont Portfolio Assessment Program as it has been implemented at the Maple Leaf Supervisory Union (M.L.S.U.) and at Maple Leaf Middle School over a two-year period, from the 1993-94 through the 1994-95 school year. The study is illustrative of the presence of the state-wide portfolio system at the local level. It also outlines the assessment program, describes the portfolio format, and discusses the impact of the program upon the M.L.S.U. and Maple Leaf Middle School.

Maple Leaf Middle School serves Maple Leaf (M.L.), a small New England town of approximately 8,000. The town's graceful churches and older buildings line the main street, and the overall ambiance reflects a small, close-knit community. However, all is not well at M.L. The town lost 550 jobs between 1992 and 1994, and is still reeling from the economic impact of that loss.

Maple Leaf Middle school, spanning grades 6, 7, and 8, enrolls about 315 students who are primarily white (98 percent), and about 33 percent receive free or reduced lunch. Because M.L. does not have a public high school, about 95 percent of M.L. students go on to the local private high school, known as the "Academy."

Data sources for this case study include several interviews and observations of professional development sessions and a review of several documents, including students' portfolios.

Participants

In Vermont, the people named in Exhibit I were interviewed.

Observations

In 1993-94, the study researchers observed a professional development session of mathematics score calibration activity organized by the area's Network Leader.

In 1994-95, the study researchers observed a Mathematics Network training session consisting of mathematics score calibration activities and a Writing Network session consisting of writing score calibration activities.

State Context

In 1988, the Vermont State Board of Education adopted the first statewide testing program in the state's history. Although initially the legislature was skeptical about the idea, the business community impressed upon the statehouse that information on student performance was essential

EXHIBIT I

Study Participants

1993-94	1994-95
• State Director of Assessment • M.L. superintendent • M.L. curriculum director • One School Board Member • M.L. Middle School principal • Three M.L. Middle School teachers • Four M.L. Middle School students • One M.L. Middle School parent • One Elementary School teacher • M.L. Network Leader • Three area teachers • Teacher Association President	• M.L. Superintendent • M.L. curriculum director • One School Board Member • M.L. Middle School Principal • Four M.L. Middle School Teachers • Two M.L. Middle School Students • A Training Director • Two 8th Grade Parents

to evaluate how the educational system was functioning and how students were doing in school. Vermont teachers, in turn, convinced the legislature that portfolios would be the best method of collecting such information; as they were averse to the possibility of a large-scale multiple-choice assessment system.[1]

In 1988, then, the Vermont Department of Education decided to develop a system of assessment consisting of portfolios and multiple-choice *uniform tests* in mathematics and in writing, with a view to expanding the program to cover other subjects. The multiple-choice strategy was included to ward off opposition to a possibly unreliable assessment system. In addition, the Board conjectured that because the uniform test would contain items from the National Assessment of Educational Progress (NAEP) assessments, comparison of Vermont students' performance with those of other states would be feasible. Although the system was not mandated, 59 of the 60 Vermont supervisory unions were participating by the 1992-93 school year.

Vermont Assessment System

As mentioned above, the state employed a strategy of combining portfolios, a standardized, *uniform* test that uses equivalent forms administered under standardized conditions for each student, and a *best piece* (usually from the portfolio) representing what the student considers to be his or her *best effort* for the year.

The multiple format assessment strategy is intended to permit comparisons across schools, districts, and supervisory unions, and advance local initiatives in assessment. In other words, the intention is the design of a system that would " . . . provide reasonable comparability across schools, but not at the cost of stifling good practice and local innovation."[2]

---

[1] Robert Rothman, Large `Faculty Meeting' Ushers in Pioneering Assessment in Vermont. *Education Week*, Vol (10), No. 6, October, 10, 1990.

Vermont Portfolio System

Because the focus of this study is innovative performance assessment, this report discusses only the portfolio component of the Vermont assessment system. The statewide portfolio strategy is the first of its kind to be established in the United States.

The stated purposes of Vermont portfolio assessments are to (a) furnish data on student performance, (b) encourage effective approaches to instruction, and (c) foster greater equity in educational opportunity. The Vermont Board of Education reasoned that the portfolio assessments would help them to achieve these goals by enabling schools to compare their scores with those of other schools, by building capacity for districts to evaluate their own mathematics and writing programs, and by helping teachers assess their instructional methods and their students' progress in mathematics and writing.[3] The assessment system, itself, is not "high stakes," as there are no sanctions and rewards linked to the assessment results.

The state compiles and submits results of the assessments to the Superintendent of each Supervisory Union and to the Principal of each participant school. The state also encourages schools to share results with the community on what the state calls "School Report Day."

The Department of Education currently spends less than $1 million of its total $199 million state education budget on its assessment system.[4]

Development of Portfolio Assessments

The portfolio assessment system was conceptualized at the state level, with the help of experts in education reform. In 1989-90, the Vermont department of education consulted with experts in education reform and education evaluation, including Grant Wiggins, Michael Fullan, Donald Graves, Dan Koretz, and the staff of the National Council of Teachers in Mathematics (NCTM).

The first pilot of the portfolio system was conducted in the 1990-91 school year. During that time, the system was still being defined through a decentralized development effort that involved local administrators and teachers.[5] A total of 138 schools across the state participated in the pilot program. The state selected 48 schools; an additional 90 volunteered to join the program.

---

[2] Koretz, D., Stecher, B., Edward, D. (July 31, 1992). The Vermont Portfolio Assessment Program: Interim Report on Implementation and Impact, 1991-1992 School Year.

[3] Ibid

[4] *News*. (Undated, sometime post November 15, 1994). Vermont Department of Education.

[5] Ibid

Full state-wide implementation commenced in 1991-92. However, RAND/CRESST (1992)[6], the evaluator of the program, described this "full-scale implementation" as " . . . a combination of a developmental effort and a pilot test, rather than as an initial implementation of a fully planned program." (p. 3).

During these initial years of development and implementation, many aspects of the portfolio assessment system were shaped through an iterative process. For example, resource materials and scoring rubrics for writing and mathematics portfolios were designed and revised by two state-sponsored teacher committees.

Portfolio Characteristics

As mentioned earlier, the state requires student portfolios in mathematics and writing. During 1991-92, 1992-93, and 1993-94, both portfolios were required of grades four and eight, but in 1994-95, the writing portfolio was shifted from the fourth to the fifth grade level — a shift intended to lighten the portfolio burden of fourth grade teachers (as both writing and mathematics portfolios were being implemented by the same teachers), and intended also to accommodate the piloting of New Standards assessments. (The writing portfolio, rather than the mathematics portfolio, was chosen for fifth grade to avoid duplication with the New Standards mathematics assessments at that grade level; the New Standards assessments Language Arts exams for grades 4, 8, and 10 and mathematics exams for grades 5, 8, and 10 were to be piloted in 35 Vermont schools.)[7]

Mathematics

The mathematic portfolio and scoring rubric are described below.

Mathematics Portfolio

Guidelines to teachers prescribe the categories and number of mathematics assignments that must be included in the fourth and eighth grade portfolios. Each portfolio must contain a set of five to seven "best pieces" selected from tasks teachers assigned to their students over the course of the academic year. Exemplars of three categories of mathematics problems — puzzles, investigations, and applications — also must be included among the five to seven "best pieces." (The two lowest-scoring pieces are dropped.) In mathematics, typical assignments are word problems that require students to apply several mathematical concepts and operations, as well as to explain their reasoning in solving the problems.

Mathematics Scoring Rubric

Each portfolio piece is graded on *communication* and *problem-solving* criteria, along a four-point scale, ranging from level one (poor) to level four (good) (see Appendix A). The *communication* criterion comprises three dimensions: (1) use of mathematical language to define the problem; (2) mathematical representation of the problem and its solution; and (3) clarity and detail (presentation) of the solution.

---

[6] Ibid

[7] *Invervals.* (October, 1994). 5(1). Vermont Department of Education.

The *problem-solving* criterion itself incorporates four dimensions: (1) understanding the problem; (2) approach to solving the problem; (3) why — reasoning in support of decisions; and (4) outcomes of activities (extension of solution to other situations). The state provides the scoring rubric to teachers at the beginning of the academic year.

Writing

The writing portfolio and scoring rubric are considered below.

Writing Portfolio

Each Portfolio for grades five and eight must contain work in certain genres, as well as a certain number of pieces of a student's writing, selected from his or her work completed over the course of the academic year. The *Writing Portfolio Table of Contents* specifies that the following writing pieces be included in the portfolio.

- A *Best Piece* of writing;
- A letter from the student to the evaluator detailing the reasons for choosing the *Best Piece*;
- A story, play, or personal narrative;
- A review of a cultural event, public exhibit, sports event, or a book, current issue, mathematics problem, or scientific phenomenon;
- Writing from any curriculum area that is not Language Arts or English (Fifth-grade, one piece; Eighth-grade, three pieces); and
- *Other Writing* (optional in 1994-95). (Poetry is optional because it is considered to be too difficult to score. It was not included in any scoring activities)

Each final paper and accompanying draft must be dated and stapled together; a student, with the help of her or his teachers, selects the sample to be contained in the final portfolio.

Writing Scoring Rubric

Each piece of writing is evaluated along the five dimensions of *Purpose*, *Organization*, *Details*, *Voice/Tone*, and *Usage/mechanics/grammar* (see Appendix B). Assessment entails using a four-point scale to judge the degree of existence, in each piece of writing, of each of the five dimensions. The scale ranges from "rarely" (1) to "extensively" (4), with the two mid-points being "sometimes" (2) and "frequently" (3).

The format of the scoring rubric was changed for use during 1994-95, as teachers in previous years had difficulty in distinguishing between "frequently" and "sometimes." Thus, for each dimension, a prompt was added between "frequently" and "sometime" to help the scorer decide whether student work should be ranked on the upper or the lower part of the scale. For example, the prompt for the *Purpose* dimension is: "*Is author's focus clear within the writing*?" A "yes" answer would prompt the scorer to consult the "frequently" or "extensively" definitions of the scale, while a "no" would prompt the scorer to consult the "sometimes" and "rarely" scale ratings.

In addition to adding the prompt, the language in the scoring rubric was modified to better describe each of the five dimensions.

In addition to a rating on each dimension, each piece of writing obtains a total score; and a holistic score is assigned to the entire portfolio. The holistic score is judgmentally based on the same scoring rubric.

Portfolio Scoring Process

As outlined below, portfolios are scored at the three levels of school, district, and state.

School-level. At the school level, teachers assess their students' homework or classwork assignments, applying state-developed scoring rubrics. They assign several tasks and then determine which tasks are to be incorporated into student portfolios.

District-level. The district-level scoring is organized by regional *Networks*, of which there are 17. The regional Networks are composed of several contiguous supervisory unions whose purpose is to organize state-sponsored professional development and to score calibration activities. The state selects a random sample (30 percent) for a score calibration session during a Network meeting of 20 to 35 teachers. All teachers in the region involved in the portfolio system are expected to attend the regional scoring session, which is led by a Network leader who usually is a teacher with extensive training and experience in the portfolio system. Teachers, in pairs, exchange and score the portfolios. If scores diverge, Network leaders assist the two teachers to negotiate the scores. The Network leaders also draw additional portfolio samples and evaluate teachers on their application of the scoring criteria. The sampled portfolios are forwarded to the state for rescoring (Vermont's Assessment Program, Undated).

State-level. During 1993-94 and 1994-95, teachers sent a state-specified random sample of portfolios to be scored by a state scoring committee consisting of teachers who are selected for scoring through an application process. (Allegedly, during 1992-93, the state asked teachers to send every fifth portfolio, but that strict direction was not followed, and poor quality portfolios often were withheld.) Network leaders collect and send their regions' sampled portfolios to a statewide meeting of selected scorers led by an Assessment Coordinator. Typically, the scoring sessions last 6 days, with scorers paid $100 a day.

Training Support

The state envisioned a system that would (a) sanction local measures of performance, (b) preserve local approaches to curriculum and instruction, and (c) still provide for a common standard of achievement across the state. Hence, much staff-development focuses on building a strong capacity to handle reform at the local level.

Throughout the year, Network Leaders offer professional development in various assessment-related activities. For example, they hold training sessions in the use of benchmarks and anchor papers and offer score calibration sessions. (One of these activities is described below). The state also provides the districts with the services of curriculum specialists. To encourage the adoption of teaching integrated curricula, the state is developing content standards for its *Common Core Framework* for curriculum and assessment, with the help of three commissions in: (a) science, mathematics, and technology; (b) arts and humanities; and (c) history and social sciences.

During 1993-94, state activities included Network sessions offered in November, January, March, and May, and a Mathematics Summer Institute in 1994. Teachers participating in the Summer Institute were requested to bring their curriculum materials and mathematics problems to share with other teachers at the institute. During 1994-95, state-sponsored activities included three Network sessions offered in November for mathematics, in December for writing, and in March and May for both mathematics and writing. (A score calibration session organized by a Network Leader is described in the District-section.) All 17 Networks offered the same professional development activities at no cost to schools.

The Department of Education, in addition to providing training support, also circulates information on, and addresses issues and questions pertaining to, the assessment system through its bimonthly newsletter, *Intervals*.

Evaluation of Portfolio Assessments

In 1988, Vermont contracted RAND/CRESST to evaluate its portfolio assessment program through 1993, and in 1993-1994, Insite, Inc., conducted the evaluation. Currently, two reports, *The Vermont Portfolio Assessment Program: Interim Report on Implementation and Impact, 1991-1992 School Year* and the *Interim Report: The Reliability of Vermont Portfolio Scores in the 1992-93 School Year*, detail interim evaluation methods and results. The reports contain a multitude of positive and negative outcomes and opinions, but only the results of the 1992-93 report and the reliability estimates of 1994-95 academic year are highlighted here.

Each year, assessment results are reported at the state, district, and school levels. Reliability of scores, validity, results, feedback from scorers, and future plans for the system are discussed in the next section.

Reliability

Between the years 1991-92 and 1992-93, *reliabilities* (as indicated by spearman correlation coefficients) based on *overall scores* (calculated from task and dimension scores) improved for mathematics portfolios, but remained low for writing portfolios (see Exhibit II).

Validity

RAND\CRESST evaluators concluded that the validity of the portfolio system could not be ascertained due to low reliabilities. For the mathematics portfolio system, the report recommended refining the scoring rubrics, placing further restrictions on the mathematics tasks for inclusion in the portfolios, providing more precise definitions of domains to be assessed, and addressing issues of task difficulty. Because teachers enjoyed a wide latitude in selecting

EXHIBIT II

Reliabilities for Mathematics and Writing

Year	Mathematics		Writing	
	Grade 4	Grade 8	Grade 4	Grade 8
1991-92*	.60	.53	.49	.60
1992-93*	.72	.79	.56	.69
1993-94**	.76	.83	.74	.69

appropriate tasks for portfolios, the quality of some tasks was questionable and comparability was compromised.

RAND/CRESST researchers concluded that the low reliability figures for the 1991-92 and 1992-93 writing portfolios represented serious design flaws, and suggested completely overhauling the system — detailing more conventional and narrower definitions of task types and designing genre-specific rubrics.

In 1993-94, the researchers from Insite, Inc. concluded that reliabilities had improved considerably, particularly for Grade 4 writing. Thus, portfolio scores could be reported at the student, school, and supervisory union levels.

The Vermont Department of Education's Director of Assessments attributes improvement in reliabilities to better training and a simpler, easier to understand writing scoring rubric.

Assessment Results

The assessment results are reported at state, supervisory union, and school levels. In 1993-94, the portfolio assessment results indicated that for both grades four and eight mathematics, over three years, the results had been stable. In Language Arts for grades four and eight, again, results were largely stable.[10]

Feedback from State Scorers

In the Fall of 1994, teachers received feedback from the state-level scorers regarding the sampled portfolios. The state scorers for the writing portfolio saw more styles of writing *and* better quality of writing in students' work in 1993-94 than in 1992-93. At the same time, they urged teachers to place more emphasis on: (1) student reflection on his or her own work; (2) developing *focus* in writing; and (3) fostering greater understanding of the different genres of writing.

---

*Source: Interim Report: The Reliability of Vermont Portfolio

**Source: Letter from Insite, Inc. to Richard P. Mills, Commissioner of Education.

[10] *Assessment Results. Writing and Mathematics. 1993-1994.* Vermont Assessment Program.

Similarly, mathematics scorers found mathematics tasks to be more appropriate for inclusion in portfolios in 1993-94 than in 1992-93. In addition, they found more evidence of student competence in problem-solving. For the coming year, the state mathematics portfolio scorers advised teachers to challenge students with a greater variety of mathematics tasks, also communicated to teachers that, although the differences in task difficulty had not been taken into account in students' scores, in future, such differences may be factored in.

Future Plans

Evaluations and scorer feedback portend further changes in the system, and the state Department of Education is beginning to address issues of mathematics task difficulty. For instance, in response to the problems identified by the RAND/CRESST evaluation, the writing rubric was modified in mid 1994 to make scoring easier and more reliable.

Furthermore, the state is developing content standards, student performance standards, and a description of essential learning experiences for the *Common Core Framework* in several content areas and in *transdisciplinary* work.[11] This on-going work is viewed as being especially important, as content and performance standards are needed to guide the assessment process. The final standards will be applicable to all students ". . . with the possible exception of an estimated 1 percent of the student population, those with severe disabilities. These exceptions should be specifically addressed in the individual students' IEPs."[12]

The standards will be developed under the umbrella of the following four categories of *Vital Results for Learning* in Vermont's *Common Core of Learning:* (1) *Communication; (2) Reasoning and Problem Solving; (3) Personal Development; and (4) Social Responsibility.*

The state also envisions extending the portfolio assessment process to other grade levels and to additional content areas, with the current plan calling for extending the mathematics portfolio to one high school grade level (which has yet to be determined), in 1995-96. In fact, over half of all high schools across the state already are piloting the mathematics portfolio in at least one class.[13] The Department of Education also hopes to add a science component to the assessment system, in either an open-ended performance or a multiple-choice format, at one middle and one high school grade level.

In 1996-97, the state envisions adding a history and social science assessment (at middle and high school levels) and, in 1997-98, an arts and a *service learning* assessment is anticipated.

---

[11] *Content Standards for the Vermont Common Core Framework for Curriculum and Assessment.* (January 6, 1995). The State Department of Education.

[12] *Content Standards for Vermont's Common Core Framework for curriculum and Assessment* (January 6, 1995). Appendix C. Vermont Department of Education.

[13] *Intervals* (January/February 1994). 4(1). Vermont Department of Education.

District Context

The Maple Leaf Supervisory Union (M.L.S.U.) operates five elementary schools and one middle school. Because Maple Leaf lacks a public high school, middle school graduates enroll in the area's private high school academies, for which the district pays the academies $6,200 per student, per year. Within its own system, the district spends around $4,500 per student, per year.

During 1993-94 and 1994-95, the M.L.S.U. confronted some daunting challenges. In 1993-94, the system faced a budget deficit of $400,000. Furthermore, according to the administrators, the system lost $400,000 in local aid for FY 1995, from its proposed total budget of $7.2 million. However, while the school system lost money, it did not lose students, and most faculty remained in the top pay-scale bracket. The 1994-95 superintendent of 20 years retired at the end of May 1994, and a new superintendent took his place in 1994-95.

In 1994-95, therefore, the new superintendent immediately delved into solving M.LS.U.'s financial problems. He met with several citizens, parent-teacher, and teacher groups, and, in the Spring of 1995, he was able to borrow $500,000 to keep M.L.S.U. financially viable. In the process, he also eliminated seven administrative and three teaching positions. During this time, he relied on the Curriculum Director to take the instructional leadership of the supervisory union.

M.L.S.U. is on the frontier of innovation, participating in the state assessment system since its inception in 1990. (The union, however, continues to utilize *Stanford Achievement Tests* every year for grades two through eight in reading and mathematics, particularly because they need this test for Chapter 1 students. The administrators have not noticed any changes in the score patterns of their students.)

Resource and Training Support

The district generally sets aside five days for staff development — three for Network sessions, and two for state-wide conferences. The three Network sessions were held during 1993-94, and three were held in 1994-95. Aside from these Network sessions, teachers also had the option of attending several state-sponsored conferences and activities. These activities included *Primary Level Writing and Mathematics: Instructional and Assessment Strategies*; a training session in *Interdisciplinary Curriculum and Assessment in the Arts and Humanities*; and an *Intensive Scoring Training* session.

During 1992-93, the district spent $1,375 on developing portfolio tasks and on scoring activities. During 1993-94, however (because teachers were scoring portfolios as an integral part of their classwork), the district spent $875.

The description of a scoring session below in Exhibit III is illustrative of the types of activities a Network meeting incorporates. In a similar Network session held in the Spring of 1995, the participants went through a comparable exercise. The post-calibration discussion revolved around how much coaching, editing, and prompting teachers should provide to students to allow for completing their portfolio work.

EXHIBIT III

Score Calibration: Observation of a 1993-94 Network Mathematics-Scoring Session[14]

In May 1994, the M.L.S.U. Network Leader, a fourth grade teacher, organized a score-calibration session for about 42 grade 4 and grade 8 mathematics teachers. Each teacher brought three mathematics portfolios, exchanged them with a partner, and rescored portfolios to determine *inter-rater consistency*. (The portfolios brought by teachers did not have to be the same ones sampled by the state.) Pairs of teachers compared student work, discussed it, and readjusted discrepant scores, taking about three hours in the process. (Later, the Network Leader will rescore 35 of the 126 portfolios brought to the meeting.) After the session, teachers debriefed one another and, as a group, listed *best* aspects of portfolios and enumerated topics they would like to see included in next year's Network meetings. The portfolios and topics are outlined below.

The *best* aspects of portfolios, in teachers' own words, were:

- *Seeing the student develop over the year*;
- *Watching students seek out and apply strategies to* [their] *work* [problems]; *and*
- [Observing] *student acceptance of multiple solutions and strategies* [for mathematics problems];

For network meetings next year, teachers want:

- *Inclusion of third and fifth grades in the portfolio system*;
- *More resources for teacher training on how to improve student success and enjoyment* [of mathematics homework];
- *Less calibration activity*;
- *More information* [on assessments] *and a sense of where to go next in the assessment* [process]; *and*
- *Information on how to better design portfolios* [in order to make them] *easier* [for the receiving teacher] *to understand.*

During the session, teachers also observed that, over the year, their students had improved the most on three dimensions of mathematics — *understanding the problem, decisions along the way to solving the problem*, and *outcomes of activities.*

The trainer said that Network leaders across the state are trying to encourage use of portfolios at all levels. She also mentioned that there is some anger about the quality of the tasks, as some districts tend to assign easy problems to their students. During the summer of 1994, thus, the state began exploring methods to judge the level of difficulty for each problem.

---

[14] The first score calibration session was held in the Spring of 1993. Score calibration sessions such as this were planned for all 17 Networks.

District Professional Development

Aside from the Network sessions, M.L.S.U. plans and conducts its own professional activities. For example, in 1993-94, the district provided training to second grade teachers on *how to use manipulatives in mathematics*. Third and fourth grade teachers received the same training in 1994-95; and teachers also receive nine free credits of college courses or professional development.

The nearby Vermont Institute of Mathematics and Science helps train teachers. In the summer of 1994, the district hosted a session entitled, *Activities that Integrate Mathematics and Science*. In preparation for this session, teachers had to devise methods to consolidate mathematics and science assignments to be included in their students' portfolios.

The Curriculum Director is especially instrumental in building an infrastructure that allows teachers to engage fully in the portfolio assessment system. The previous superintendent described the Director as "The spearhead of the activities of state assessment." The Curriculum Director provides, for one thing, release time so that all teachers can collaborate on scoring and on the design of portfolio tasks. For example, during release time in 1993-94, fourth grade teachers worked together "across buildings" to develop a battery of mathematics problems. The Director also required teachers from grades one through six to keep student portfolios.

During 1994-95, the Director continued to support the use of portfolios at all grade levels. She provided fifth grade teachers with seven half-days of release time in September and in February, specifically to enable them to plan and score their portfolios. For these periods of time, she hired three people to conduct special technology and design projects with fifth grade students. In addition, second and third grade teachers received four days of release time in the Fall of 1994; and the Director also coordinated a session on how to design rubrics, because she feels that such sessions help teachers to transmit performance expectations to their students, because ". . . rubrics have a powerful effect on student learning when expectations are defined."

The expenses associated with the use of portfolios is related to teacher release time for Network meetings. The district pays $50.00 a day for substitute teachers, while the state generally pays for supplies and materials. According to the Curriculum Director, substitute teachers are not hard to find, although, according to the 8th grade Language Arts teacher and the M.L. Middle school Principal, *quality* substitutes are difficult to find.

The district is interested, too, in establishing its own curricular frameworks, based on the common framework the state develops. At the moment, therefore, the district is taking a "wait-and-see" approach to this reform. In the meantime, the Curriculum Director organized several workshops around the topic of curriculum. For example, two experts from the Vermont Institute of Mathematics and Science held workshops exploring the connections between content standards and scoring rubrics. The Curriculum Director also is collaborating with the College Board and the North East Regional Lab on understanding and defining curriculum focus in the areas of science, mathematics, and technology. She, herself, is involved with the Association for Supervision and Curriculum Development.

Impact of Vermont's Portfolios

District officials' and teachers' opinions and judgments are suggestive of the intended and unintended effects of the portfolio system on M.L.S.U., as can be seen in the following discussion.

Impact on Administrators

In both 1993-94 and 1994-95, district officials included in this study were enthusiastic about the portfolio system, but they were unsure about how to use the data they accumulated. The Curriculum Director stressed that they were ". . . careful about interpretation. Issues around reliability and validity are challenging." She would like to ensure reliability and validity of the system, so that the process can continue. However, the true contribution of the system, she emphasized, is in its validity for instructional, rather than assessment, purposes, and views the system as the predominant strategy for educational reform. In fact, in her 25+ years of experience as an educator, the Director declared that, ". . . nothing has changed instruction as dramatically as the Vermont system because of the accountability issue — to teachers, personally, it is high stakes." Overall, she said, "the process is enlightening," and both superintendents shared her enthusiasm.

The new superintendent was especially pleased with the idea of portfolios, because he felt that portfolios "force" teachers to ". . .talk across content areas." He noted, however, that reliability remains a problem with the system, but parents want to be able to compare M.L.S.U. with other supervisory unions. He is not sure, therefore, that the community will buy into the system, if it cannot be used for accountability purposes. Too, he is concerned about the system in terms of time-consumption.

In addition, both he and the Curriculum Director concurred that in order for the portfolio assessments to continue to function as informative pedagogical tools, the curriculum must be better defined. The Curriculum Director currently is serving on two state-level committees, *Curricular Frameworks* and *Essential Learning Experiences*, to help define the *Core Curriculum*, which is to form the bedrock of the local curriculum. Reform of the curriculum is, then, just beginning to catch up with reform of the assessment system.

Impact on District Teachers

In 1993-94, according to the Curriculum Director, some Language Arts and mathematics fourth and eighth grade teachers expressed that the portfolio system was unjust, because the onus of ensuring portfolio completion fell on their shoulders. These teachers expressed that the responsibility for portfolios should be shared with the lower grades and with other teachers. The Director said that the system was perceived as being especially difficult for the Language Arts teachers, because they had to contend with the additional demand of obtaining student writings across the curriculum. In 1994-95, the burden on the fourth grade teachers was lessened somewhat, because the Language Arts portfolio was moved up to fifth grade.

As previously discussed, students are required to submit at least three pieces of writing from subject areas other than Language Arts, and Language Arts teachers do not want to include unacceptable pieces from other classes in their students' portfolios. Some Language Arts teachers, therefore, have devised a solution to this problem by sharing writing criteria with other teachers.

Another ongoing debate in the region concerns whether or not extensive writing is an effective means to learn content areas (e.g., science and mathematics). She suspects that mathematics teachers are skeptical of the system, not because they are skeptical of portfolios, per se, but because they question the primacy of the NCTM standards in the scoring rubrics. The Director indicated that many teachers felt that the mathematics portfolio should have been moved up from the fourth to the fifth grade. These teachers believe that some of the demands of the mathematics portfolios, such as the *So What* criteria of the *Problem-solving* dimension, are not developmentally appropriate for fourth grade students.

The Curriculum Director also said that many teachers have realized that everything from a unit cannot be included in performance based activities, and that certain features of the units have to be excluded.

The union. To date, the teacher's union, which, according to two-time study participants, supports good teaching practices, has presented no organized opposition to the assessments. In 1993-94, however, many teachers were suspicious of the state's sampling procedures. Even though the state randomly samples student portfolios, somehow, some teachers said the state had managed to select portfolios of poor quality.

In 1994-95, no such sentiment was evident, and the representatives said that the benefits of the portfolio system outweighed the costs. The benefits of being able to assess the "whole child on an individual basis" and of having an assessment tool that could inform curriculum were viewed as outweighing the costs of subjectivity in scoring. Nonetheless, representatives felt that the comparability across students is an important issue that must be solved.

A second important piece of the assessment process is the *Common Core* framework. The Union representatives said that the framework must be better defined in order to support good assessment practices.

Fourth grade teacher. The fourth grade teacher who participated in this study in 1993-94 observed that her instructional approach had changed, but that the change did not derive impetus from the use of portfolios. She required students to self-assess themselves, using the scoring rubrics, and found that their scores were generally not discrepant from her own.

Fifth grade teacher. In 1994-95, the fifth grade teacher who participated in this study was from the neighboring supervisory union and was conducting a writing Network session at Maple Leaf. She said that the critical need for fifth grade teachers at this point is the availability of benchmarks. For example, because fifth grade teachers are new to the writing portfolios, they are not certain of what exemplifies *extensively* on the scoring rubric. Fifth grade teachers received their first Network training in December, and the fifth grade benchmarks were received only in mid-March. These teachers also are just beginning to use a common portfolio terminology; they are still determining to what extent they should help their students complete portfolio assignments. This entails, for example, thinking about how much to edit, how much to help students with their revisions, and how many times to ask students to revise their work.

This teacher expressed the worry that poor school systems that are likely to be at the bottom of the state due to insufficient resources also are most likely to feel accountability pressures. Given this scenario, she said that she would not be surprised if teachers provide "extensive help" to their students.

School Context: Maple Leaf Middle School

The concern at Maple Leaf Middle School is to provide the kind of nurturing environment that enhances academic and social outcomes for Maple Leaf students, not only in the Middle School, but also in the private Academy. However, the Principal and teachers said that now, more than ever before, students are entering school with tenuous academic and social preparation, and these exigencies place even greater demands on teachers. Thus, while contending with demands of the relatively new portfolio system, teachers also must prepare students for the Academy, as well as provide "social services," such as counseling, to their students. Despite such a heavy burden, the Maple Leaf Middle School community has met the challenges with admirable tenacity.

Training Support

During 1993-94, teachers benefited primarily from the state- and district-supported professional development sessions, but they also maintained an informal network of support among themselves. However, several teachers mentioned that they desired an even greater amount of in-house support for developing and scoring portfolios.

In 1994-95, the eighth grade Language Arts teacher took the initiative to provide such support, in order that other teachers could start using the writing process. This teacher collaborates with other teachers on a regular basis to help them institute the writing process in their classes. For example, she helped the social studies teacher develop social studies assignments and a "checklist" for scoring those assignments (see Appendix C).

Impact of Vermont's Portfolios

Impact on Administrator

M.L.S.U.'s principal of seven years solidly supports the portfolio system. Retiring at the end of 1994-95, he maintains that portfolios promote thinking and, hence, tried ". . . to convince the Language Arts teachers that it is not only their burden." The principal wants all teachers to take "ownership of portfolios." Issues of reliability and validity do not concern him at the moment, because " . . . they [the teachers] will do it [mange the portfolio system] better as they go along." True to his convictions, the principal hired a computer aide who helps students use the computer, which is intended to provide assistance with the Language Arts portfolios.

While the Principal exudes enthusiasm about the effect of the assessments, he harbors concerns regarding the logistics of implementation. He is worried about "time and money." Because several teachers attend training sessions at the same time, several times a week, payment for substitute teachers is a real burden. In 1994-95, he also expressed concerns about the burdens of the system imposed on the Language Arts teacher, as he felt that it is not only the writing teacher who ought to be involved in coordinating the writing portfolios; the involvement of other teachers is essential.

The principal is uncertain about how the results are presented to the public, and about whether or not state grants can be linked to assessment results in the future. Despite these uncertainties, and the anxieties they generate, he feels that the assessments are worthwhile, because ". . . people have hope in the system and they [the people] feel that system is important," he said.

Impact on Teachers

M.L.S.U. participates formally in the Vermont portfolio system through its one mathematics teacher and two Language Arts teachers. Informally, other teachers are "trying their hands" at the system; and those teachers expressed a variety of opinions about the portfolio system and are integrating it into their classrooms in a variety of idiosyncratic ways. The seventh grade mathematics teacher, too, has adopted the portfolio system.

Eighth grade mathematics teacher. The eighth grade mathematics teacher harbors conflicted opinions about the mathematics portfolio. His experiences and his thoughts about the portfolios did not change from one year to the next. On the one hand, he is uncertain about the scoring procedures (i.e., reliability), and uncertain whether the portfolio process will improve the mathematics curriculum. For various reasons, he does not believe the system is valid and is worried that students will lose basic skills, such as the ability to calculate fractions and to remember multiplication tables. "I don't think mathematics should be exclusively writing," he said, and also expressed the worry that eventually the system will be used to judge teachers.

On the other hand, this teacher is encouraged by the evidence that portfolios advance the application of mathematical language and concepts. (He, like others previously mentioned, shares scoring rubrics with his students before task assignments.) Too, the system has enabled him to better understand a child with limited facility in mathematics. Thus, he uses only portfolio completion (and not the quality of work in the portfolios) for a portion of the class grade. For him, the mathematics portfolio is an "add-on," not an integral feature of his classroom activities.

Seventh grade mathematics teacher. In 1994-95, the seventh grade mathematics teacher's overall responses diverged from those of the eighth grade teacher in that she was considerably more responsive to what the portfolio system offered. Her view became that seventh grade was the "practice year" for being able to assemble portfolios during eighth grade. Therefore, she had incorporated portfolios into her daily teaching routine. However, she, too, ascribed to the conviction that the use of the "problem-stressing aspect" of mathematics is not enough to convey mathematical concepts and skills.

This teacher noted that she spent a fair amount of time teaching students English writing skills, rather than mathematics skills, per se. Usually, she required her students to complete a problem every two weeks, from which they picked three or four "good pieces" to include in their portfolios; and she generally utilized problems from NCTM-based textbooks, but felt that the scoring process relied too much on subjectivity and, therefore, hampered comparison *across* students.

In 1994-95, she continued to use the portfolios, but with more misgivings. Although she believes that all rubric dimensions are valid, she is concerned that, in fact, some children are very good in mathematical skills and thinking, but cannot write in English, and, therefore, develop negative attitudes toward mathematics. During this year, she also began to feel that teaching her students communication skills was taking time away from teaching mathematical skills. She did not devote as much time to teaching certain units (e.g., integers, rational numbers, geometry, fractions, probability) as she thinks is required for students to truly grasp the concepts. She tried to integrate her class with Language Arts, but coordination was difficult. She now prefers the idea of less writing in mathematics and advocates teaching basic skills without necessarily embedding them in applied problems.

Eighth grade Language Arts teachers. Both eighth grade Language Arts teachers expressed positive opinions about the writing portfolio system. They believe that, although the scoring standards will never emulate the precision obtainable with standardized tests, as one said, "they are adequate for now."

During 1993-94, only one Language Arts teacher participated in this study. At that time, she said that her pedagogical approach centered on guiding students to think and to share thoughts through peer-conferences, rather than on the syntactical and grammatical aspects of writing. She was enthusiastic about stressing the *voice* dimension in writing, as it had a particularly positive impact on her female students — she found that their writing became "natural and reflective of their experiences as females." Her students produced 8 to 12 pieces of writing, each of which was 2 pages long; and she based her students' grades on those pieces.

By 1994-95, this teacher said that she had integrated the writing process even more firmly into her classes. She held a 45 minute writing workshop for her students on a weekly basis to encourage them to adopt the writing process, and she focused on teaching students skills in using the writing scoring rubric to judge their own strengths and weaknesses. Interestingly, but not surprisingly, she also began to pay much more attention to the grammatical and syntactical aspects of writing.

This teacher also insisted that portfolio pieces deriving from other classes show evidence of draft work, peer-conferencing, teacher-student conferencing, and that the work be scored by those teachers. She believes that, in fact, because of such insistence on the writing process, other teachers are actively stressing written communication in their classes.

In her estimation, use of the portfolio system requires a strong teacher with good organizational skills, initiative, and commitment. However, she also fervently believes that the approach itself can "transform" teachers. "It has made me a better teacher," she said. In fact, as a result of being in charge of the writing portfolio, both Language Arts teachers, became mentors to other teachers, showing them how to use the writing process and performance-based assignments.

Both teachers found the writing scoring rubric easier to use in 1994-95, essentially because of the prompts that were added to it. However, they are still not that concerned about absolute precision in scoring. "I don't care whether it is a good assessment. I think its validity is in the classroom. They [students] are learning, performing, producing, and if calibration isn't accurate, I don't care. The writing process is important," said one teacher.

During 1994-95, these teachers also learned to score portfolio pieces on an on-going basis, making the task a little less time-consuming than it was during the 1993-94 school year. One of the teachers said that he actually uses a "simplified Likert" rubric on top of each piece of student work, to expedite the scoring process.

Both teachers were, however, concerned with articulation of standards and curriculum with the Academy. Allegedly, the Academy teachers feel that the M.L. Middle school graduates do not possess sufficient mastery of grammar, syntax, and spelling. During 1994-95, thus, M.L teachers began stressing these aspects of writing in their classes Prior to using the portfolios, one teacher had placed emphasis on grammar; then, she paid much more attention to the "voice" and creative aspects of writing; and now she is placing equal emphasis on both.

Impact on Students

Student responses to the portfolio assessments can be characterized as a blend of both enjoyment and indifference. The four students included in this study during 1993-94 did not think the mathematics portfolios were much "fun." However, they did mention that some mathematics problems were challenging and that those problems prompted them to plan and to reflect upon how to write. (The eighth grade mathematics teacher said that students who enjoyed writing liked the mathematics portfolios, but those who did not enjoy writing did not like the portfolios.)

The students exhibited a different attitude toward the Language Arts portfolios. They spoke of writing more than they ever had before, and said they found the challenge stimulating. "You have to think more about it to get a better grade," they said. The Language Arts teacher observed that her students at first felt awkward about writing about matters of the "heart." However, later they felt "vested" in their writing and were much more motivated to complete assignments. Maple Leaf teachers mentioned, however, that some children feel disadvantaged when they graduate and enroll in the Academy, because they lack the content skills their peers from other schools possess. Some teachers also mentioned that "low-ability students" were frequently more successful in school as a result of the portfolios. However, other teachers said that those same students often experienced difficulty with portfolio tasks.

The two student participants during 1994-95 had reactions similar to the 1993-94 students, saying that the mathematics problems they have to do for their mathematics portfolios are difficult, "more in-depth," "complex," and require a lot of writing. However, they had to complete those problems at home (not in the classroom as they had done when they were in 7th grade), and their scores on the problems are not factored into their grades (as they were in 7th grade). Their grades were based on in-class tests that also present challenging word-problems, but they did not have to write more than a sentence or two to explain their solutions.

Both students indicated that they typically do not use the mathematics scoring rubrics to evaluate their own work. Although the teacher reviews the portfolio problems in class, he typically stresses "the mathematics part, not the writing part," and asks them to go to the Language Arts teacher to improve their writing skills in mathematics.

Both students indicated that they enjoy writing for the writing portfolios. They said that the experience was very different from last year, when they did "lots of grammar, and opinion stories, . . . It wasn't as much free-writing." In 8th grade, they added, they learned to do research papers and different genres of writing.

Both students also claimed that looking over their portfolio pieces helps them to examine their proficiency in Language Arts. They are required to use the scoring rubric to set standards in writing skills and self-evaluate their progress vis-a-vis these standards, ergo they apply the scoring rubric to their work, assess their weak points, and try to make improvements. "I used to have a problem with detail in my stories. I use the rubric for [paying attention to] detail in the story," said one student.

These students also mentioned a fledgling use of the writing process in classes other than mathematics and Language Arts. They said that their science and social science teachers are attempting to use more research and experiment-based assignments that require written reports and scoring rubrics. Both said they "enjoyed" doing such assignments.

Overall, during 1994-95, the effects of portfolios on student achievement and motivation appear to be the same as in 1993-94. One Language Arts teacher noted, however, that writing has become a habit with students, and that they no longer think of it as quite so burdensome.

Impact on Special Education Children

Some degree of confusion exists regarding portfolio requirements for special education children. Although the Department of Education's official policy states that all portfolios are required for *all* children, with IEPs included for special education children, some teachers mentioned that IEPs are included in the portfolios, while others said that portfolios are not marked in any way for special education children. Some teachers specify in IEPs that portfolios are not required, despite the fact that some of these children "shine" in portfolios.

Impact on Curriculum and Instruction

The consensus at M.L. Middle School is that teaching practices have changed, but that changes cannot be attributed entirely to the portfolio system. The eighth grade mathematics teacher indicated that his curriculum has not been affected much (and instruction only slightly), but he is now using more manipulatives in mathematics. The seventh grade mathematics teacher, however, noted that she had modified her teaching strategies as a result of the portfolio system. In both years, she noted that she engaged her students in "problem-solving" through the use of mathematical concepts. However, in 1994-95, she also noted that she did not spend as much time on certain units of the curriculum (e.g., integers and rational numbers) as she had wanted to spend, because the portfolio process is very time consuming.

Both Language Arts teachers' strategies, however, have been deeply affected by the use of the portfolio. They emphasize different aspects of writing as a result of using the scoring rubric and portfolio requirements, and also use the rubric to set performance expectations for their students. They distribute writing benchmark pieces to their students, use the writing process, use the scoring rubric to design student assignments, and have become involved in writing across different content areas. In short, these teachers have fully integrated the portfolio process into their daily teaching practices.

Impact on Parents and Community

Parents. In both years, parents interviewed did not have clear and sufficient information about the portfolio system. In 1993-94, the parent participant was the president of the Parent-Teacher group at Maple Leaf; and her daughter was in the eighth grade. This parent remarked that she did not understand entirely the purposes of the portfolio system, and neither did other parents. "I'm pretty involved with kids," she said, "and if I'm hazy on portfolios, then most everyone is."

Because of lack of information, the parent professed neutrality toward the system. "I guess I don't really have a strong opinion [about the portfolio system]." She felt that portfolios were acceptable so long as they helped with assessing student educational progress. However, she tended to rely on teacher judgement and the district-wide Stanford Achievement Tests (SATs) to obtain information on her child's progress in school. (The board, too, gets a full report on the SATs.)

Nonetheless, this parent viewed her daughter's enjoyment of the portfolios and knowledge of the connectedness of mathematics concepts as a ". . . good outcome of the portfolio system." Overall, her inclination was to trust the school system, and she felt that her daughter was "doing well" in school. Her major worry, aside from the portfolio system, was that many children were bringing social problems associated with poverty and drug use into the school — and that education was becoming "diluted" due to those "social trends."

In 1994-95, parent participants' reactions were considerably more positive, but reflected that they were only slightly more aware of the portfolio system than they were the previous year. One parent said that the portfolios "made sense" to him, because he values "penmanship" and because he believes portfolios to be a better system of evaluating student work than what was in place. The other parent likes the integration of mathematics and writing. Both, however, professed to knowing very little about the portfolio purposes and processes. What is important to both of them, though, is to see their children happy and engaged in learning.

Both parents said that they see report cards and graded materials much more often than portfolio products; and they rely on the grading system to monitor their children's progress in school.

One parent said that although he ". . . falls into the trap of student scores," he does not like standardized, multiple-choice tests because they do not measure students' full talents and skills. He said that most parents are open to the idea of performance assessments. The other parent, however, would like to see ". . . standardized testing of some sort" used in the school system. The need for the school system to involve parents in the portfolio assessment process was, thus, quite evident in both years.

Community. Thus far, the Maple Leaf community has not presented an orchestrated response to the portfolio system, and may not be entirely aware of the system's existence. In 1993-94, the board member intimated that parents are much more interested in their children's progress than in comparing their children's scores with those of students from other districts. However, he did say that some people are questioning the state's motives with regard to what it will actually do with the data; and some do not understand why only fourth and eighth graders were chosen for the assessments.

For 1994-95, however, the story had changed somewhat. The school board member expressed concerns over the "subjectivity" of the scoring rubric, and wants some clear "measurements of our results." He also believes that not enough information is available with regard to what impact portfolios have on instruction, curriculum, and student learning. He said that he does not approve of the *Common Core* as it "underemphasizes content;" nor does he like "touchy-feely" mathematics, because it underemphasizes certain rote knowledge, such as multiplication tables. Overall, he is worried that students are not leaving the school system with ". . . a reservoir of basic information and education that they didn't have when they first started." Of the five Local School Board members, he said, one other has views similar to his own. However, he is not certain about the other three people's ideas.

The school-level system for informing the community remains to be operationalized at Maple Leaf Middle school. Although two teachers held "portfolio nights" with some parents, the school did not hold a "School Report Day" in 1993-94, and none is planned for 1994-95.

Future Plans

The district has a *documented curriculum*, but all areas will be revised fully after the state defines the *Common Core Curriculum*. The Curriculum Director also had hoped to incorporate the portfolio assessment system at all grade levels by 1994-95. That process, however, is still continuing. She said it would be "difficult" to establish a performance baseline and to use the assessments for "high stakes decisions," because learning opportunities — such as access to technology — for students "are not uniform." In the arena of professional development, the Curriculum Director and the new superintendent plan to organize activities around integrating curriculum and assessment.

The new superintendent also plans to reorganize Maple Leaf Middle School into three teacher-student teams. He wants to encourage team teaching and block scheduling, and he wants to introduce more technology into the school.

Conclusions

The findings highlighted at the state level are the ones being verbalized, with earnestness and reflection, by teachers at Maple Leaf. Teachers and others at the local level are not entirely clear about the content validity, reliability, and quality of the portfolio program. Articulation across grades and classes and the connection between content and assessments are the other large issues facing Maple Leaf. The concern regarding the connection between content and assessments is especially strong for the mathematics portfolios. Nonetheless, Maple Leaf administrators have embraced the portfolio system and have given it their full support.

APPENDIX A

Mathematics Scoring Criteria

MATHEMATICS COMMUNICATION CRITERIA

<table>
<tr><th>Vermont Portfolio Scoring Guide: Mathematics Rev 3/5/93</th><th>Level 1</th><th>Level 2</th><th>Level 3</th><th>Level 4</th></tr>
<tr><td rowspan="2">C1 Mathematical Language</td><td><i>...didn't use any math vocabulary, equations, or notations or used them incorrectly.</i></td><td><i>...used basic math words or basic notation accurately.</i></td><td><i>...went beyond occasional use of basic math language and used the language correctly.</i></td><td><i>...relied heavily on sophisticated math language to communicate the solution.</i></td></tr>
<tr><td colspan="4">Basic Mathematical Language is limited to the math words and symbols that are commonly used at the student's grade level. For example, words like multiply and subtract, or notation like + and =. Don't consider math language that is included in the problem statement. Simply repeating words in the problem earns only Level 1.
Sophisticated Mathematical Language includes language not commonly used at this student's grade level. For example, words like exponent or sequence, or notation like $x < 6$ or $\bar{2}$.
Special Considerations: One mistake in accuracy will not drop a score down to a Level 1. Similarly, use of one math term rarely merits scoring a Level 2.</td></tr>
<tr><td rowspan="2">C2 Mathematical Representation</td><td><i>...didn't use any graphs, tables, charts, models, diagrams or drawings to communicate the solution.</i></td><td><i>...attempted to use appropriate representation.</i></td><td><i>...used appropriate math representation accurately and appropriately.</i></td><td><i>...used sophisticated graphs, tables, charts, models diagrams, or drawings to communicate the solution.</i></td></tr>
<tr><td colspan="4">An Appropriate Representation is one that is related to the problem. Using the representation appropriately is executing the representation properly.
Mathematical Representations include graphs, charts, tables, models, diagrams, and equations that are linked to representations. Completing a structured chart at Grade 4 earns Level 2; at Grade 8 earns Level 1.
Accurate graphs include: 1) labeled and correctly scaled axes, 2) appropriate titles, and 3) keys, when necessary.
Sophisticated representations include perceptive representations; combinations of many graphs, charts, and tables to organize, display, and link data; and representations that were relied upon to obtain a solution.
Special Considerations: For a representation to be related it must reflect the problem.</td></tr>
<tr><td rowspan="2">C3 Presentation</td><td><i>...response is unclear.</i></td><td><i>...response contains some clear parts.</i></td><td><i>...if others read this response, they would have to fill in some details to understand the solution.</i></td><td><i>...response is well organized and detailed.</i></td></tr>
<tr><td colspan="4">Unclear suggests the reader has little or no idea what was done to solve the problem.
Some clear parts suggests the reader understood some of the work but has so many questions about what the student did that the reader is uncertain what was done to solve the problem.
Fill in some details means that although most of the solution is organized, it may be missing some details or it is detailed but lacks organization, and the reader is required to fill in.
Well organized pieces of work have all the parts connected to each other (e.g., if there are graphs and tables, there is an explanation of their part in the solution).</td></tr>
</table>

MATHEMATICS PROBLEM SOLVING CRITERIA

Vermont Portfolio Scoring Guide: Mathematics Rev. 6/3/93	Level 1	Level 2	Level 3	Level 4
PS1 Understanding the Problem	*...didn't understand enough to get started or make progress.*	*...understood enough to solve part of the problem or to get part of a solution*	*...understood the problem.*	*...identified special factors that influenced the approach before starting the problem.*
	Part of a Problem: Some problems are multi-step. If all the parts of the problem are not addressed, then all of the problem was not understood. **Solution:** A solution includes all of the work that was done to complete the problem, an explanation of the decisions made along the way, and an answer. **Special Factors:** Special factors are factors that might affect the outcome of the problem. **Special Considerations:** A Level 4 for this criterion is dependent on identifying factors to be considered before starting the problem.			
PS2 How You Solved the Problem	*...approach didn't work.*	*...approach would only lead to solving part of the problem.*	*...approach would work for the problem.*	*...approach was efficient or sophisticated.*
	Approach: The strategy or skill used to solve the problem. **Would:** An approach that would work for a problem even if computation errors or an incomplete response prevented a solution is credited as a Level 3 **Efficient:** Efficiency is determined by the directness of the approach. Use of an algorithm to solve a problem suggests this was just an application of knowledge, not a real problem. If finding the least common multiple, an approach which lists all multiples of a number as compared to the use of prime factorization is not as efficient. The second case might be sophisticated because of its efficiency. **Sophisticated:** A sophisticated approach is one that is not common for students this age. **Special Considerations:** A piece scored at a Level 1 or 2 on PS1 can not score more than a Level 2 on PS2.			
PS3 Why - Decisions Along the Way	*...no reasoning is evident from the work or reasoning is incorrect.*	*...only partly correct reasoning or correct reasoning was for only part of the problem.*	*..didn't clearly explain reasons for decisions, but work suggests correct reasoning used throughout the problem.*	*...clearly explained the reasons for the correct decisions made throughout the problem.*
	Suggests: Work suggests there is reasoning if: there is a change in approach, but no reason given for the change; there is more than one approach to the problem but no comparisons were made to show this was done as verification; work of other approaches is given but without explanation of their part in problem solving. **Clearly Explained:** Clear reasoning can be seen in the following ways: a written explanation of the decisions made along the way; a written justification for why a path was followed; multiple approaches with comparisons and verifications; and commentary on the approach that shows evidence of thinking.			
PS4 So What - Outcomes of Activities	*...solved the problem and stopped.*	*...solved the problem and made comments about something in the solution.*	*...solved the problem and connected the solution to other math OR described a use for what was learned in the "real world."*	*...solved the problem and made a general rule about the solution or extended the solution to a more complicated situation.*
	Connections can be: between mathematical ideas; between problems; to other classes or content areas; to other cases. **General Rule:** a rule that can be used no matter what the numbers in the problem are. A general rule need not be an algebraic rule, it can also be a generalization of the problem to a more complicated situation. **Prompted Response:** General prompts (e.g., Can you think when you might use this? Do you notice anything interesting in your solution?) are OK for a teacher to give. Specific prompts (e.g., What does this problem have to do with factors? How is this similar to pricing items at a grocery store?) limit scoring to a Level 2. **Special Consideration:** For this criterion, the Levels are independent of each other. For example, a student could score a 4 without a comment or description of real world use. Score Level 4 if a generalization was made at any point in the problem, whether a requirement of the problem or not, as long as an explanation showing understanding or derivation of the generalization is included.			

APPENDIX B

Writing Analytic Assessment Guide

Vermont Writing Assessment

Analytic Assessment Guide

	Purpose	Organization	Details	Voice or Tone	Grammar/Usage/Mechanics
In assessing, consider...	...how adequately intent and focus are established and maintained (success in this criterion should not depend on the reader's knowledge of the writing assignment: the writing should stand on its own)	...coherence: ...whether ideas or information are in logical sequence or move the piece forward ...whether sentences and images are clearly related to each other (Indenting paragraphs is a matter of Grammar/Usage/Mechanics)	...whether details develop ideas or information ...whether details elaborate or clarify the content of the writing with images, careful explanation, effective dialogue, parenthetical expressions, stage directions, etc.	...whether the writing displays a natural style, appropriate to the narrator ...*or* whether the tone of the writing is appropriate to its content	...the conventions of writing, including: •Grammar (e.g. sentence structure, syntax) •Usage (e.g. agreement and word choice) •Mechanics (e.g. spelling, capitalization, punctuation)
Ask how consistently, relative to length and complexity...	intent is established and maintained within a given piece of writing	the writing demonstrates coherence	details contribute to development of ideas and information, evoke images or otherwise elaborate or clarify the content of the writing	an appropriate voice or tone is established and maintained	As appropriate to grade level, command of conventions is evident, through correct English or intentional, effective departure from conventions
Extensively	Establishes and maintains a clear purpose and focus.	Organized from beginning to end, logical progression of ideas, fluent and coherent.	Details are pertinent, vivid or explicit and provide ideas/information in depth.	Distinctive personal expression or distinctive tone enhances the writing.	Few or no errors present; *or* departures from convention appear intentional and are effective.
Frequently	Establishes a purpose and focus.	Organization moves writing forward with few lapses in unity or coherence.	Details develop ideas/information; *or* details are elaborated.	Establishes personal expression or effective tone.	Some errors or patterns of errors are present.
	Yes ***Is author's focus clear within the writing?*** No	Yes ***Does the organization move the writing forward?*** No	Yes ***Do details enhance and/or clarify the writing?*** No	Yes ***Can you hear the writer? Or, is the tone effective?*** No	Yes ***Does writing show grade-appropriate command of G/U/M?*** No
Sometimes	Attempts to establish a purpose; focus of writing is not fully clear.	Lapse(s) in organization affect unity or coherence.	Details lack elaboration, merely listed or unnecessarily repetitious.	Attempts personal expression or appropriate tone.	Numerous errors are apparent and may distract the reader.
Rarely	Purpose and focus not apparent.	Serious errors in organization make writing difficult to follow.	Details are minimal, inappropriate, or random.	Personal expression or appropriate tone not evident.	Errors interfere with understanding.

NONSCORABLE

- *is illegible: i.e., includes so many indecipherable words that no sense can be made of the writing, or*
- *is incoherent: i.e., words are legible but syntax is so garbled that response makes no sense, or*
- *is a blank piece of paper*
- *For Portfolio: Does not have required minimum contents*

©1994 Vermont Department of Education

APPENDIX C

Language Arts and Social Studies Rubric

8TH GRADE RESEARCH PAPER

RUBRIC FOR LANGUAGE ARTS AND SOCIAL STUDIES

1. EVIDENCE OF THE WRITING PROCESS
2. GOOD OPENING
3. CLEAR THESIS STATEMENT
4. CLEAR PLAN OF DEVELOPMENT
5. CONTAIN AT LEAST ONE APPROPRIATELY DOCUMENTED DIRECT QUOTE
6. SAME FOR INDIRECT QUOTE
7. END NOTES [PROPER FORM]
8. BIBLIOGRAPHY [PROPER FORM]
9. SELF ASSESSMENT – RUBRIC FROM VT STATE PORTFOLIO
10. CONTAIN SOME PERSONAL OPINION
11. COVER ONE OR MORE OF THE FOCUS QUESTIONS
12. TYPED – COMPUTER TIME WILL BE MADE AVAILABLE

CRANDALL

Oregon Assessment Reform:
Crandall High School
March 11-12, 1994
April 17-18, 1995

OREGON ASSESSMENT REFORM: CRANDALL HIGH SCHOOL

Introduction

This study summarizes Oregon's student assessment reform as it is being formulated and implemented at District A, with a special focus on Crandall High School. Because Oregon's assessment reform is an integral component of its overall education reform strategy, this paper begins by summarizing the history and goals of state education reform, and then turns to the district's and to the school's involvement in and reactions to the state's assessment reform component. The study describes the district's involvement over two academic years, 1993-94, and 1994-95. However, Crandall's history of involvement covers only 1993-94, as it withdrew from full-scale participation in the pilot reform efforts and, therefore, from this study. The Principal indicated that because Crandall was no longer participating in assessment task development, Crandall teachers did not want to participate in this study.

Information for this case-study is drawn from interviews with a number of individuals at the state, district, and school levels, from documents, and from observations of classes and performance development sessions.

Participants

Roles of the interviewees are shown in Exhibit I.

Observations

In 1993-94, observations of a professional development session devoted to the use of a scoring rubric and an in-class student performance observation furnished further information for this case-study. In 1994-95, the study researchers observed two professional development sessions devoted to the design and scoring of portfolios and one class period.

EXHIBIT I

Study Participants

1993-94	1994-95
• Assistant Superintendent for Assessment and Technology, Oregon Department of Education • District Curriculum Director • District Superintendent • High School Principal • High School Assistant Principal • Nine high school teachers • Two high school students • One Parent • One School Board member	• District Superintendent • District Curriculum Director • Two elementary school principals • Eight elementary school teachers • One middle school teacher • One middle school parent

State Context

Oregon's assessment reform can be understood only within the context of Oregon's educational reform agenda, which began with the passage of the *Educational Act for the 21st Century.*

Oregon's Education Reform

In 1991, the legislative Assembly in Oregon enacted the Oregon Educational Act for the 21st Century (House Bill 3565), the primary purpose of which was to usher in a new, state-wide high-standard educational system. The Act was passed to ". . . achieve the state's goals of the best educated citizens in the nation by the year 2000 and a work force equal to any in the world by the year 2010" (Section 2 (3)). Influenced by the National Center of Education and the Economy's report, *High Skills, Low Wages,* those who passed the Act hoped, in particular, to have an impact upon work-related student outcomes.

The passage of the Act resulted in the articulation of an outcomes-based educational system with criteria enumerated for the *Certificate of Initial Mastery* (CIM), to be awarded at about the end of grade 10, and a *Certificate of Advanced Mastery* (CAM), to be awarded at about the end of grade 12. Award of certificates were envisioned to depend upon students demonstrating certain abilities, not upon students having spent a certain amount of time in the classroom, as the state had hoped to abolish grade-levels and the high school diploma itself. (Benchmarks were to be established at earlier levels, too.) The first CIM was be awarded to the 1994-95 eighth graders in 1997. The general CIM outcomes are shown in Exhibit II.

Oregon's Educational Act for the 21st Century proved to be controversial, and, in March of 1995, the House passed House Bill 2991 to amend the Act. House Bill 2991 addressed and modified several controversial aspects of the original Act.[1] Major provisions of the bill included, for example, addition of references to content and academic standards, elimination of the *Certificate of Initial Mastery* and *Certificate of Advanced Mastery,* linkage of assessments to academic content, and setting of standards at the state level. In June, 1995, the Governor signed the Bill, which eliminates all CIM outcomes discussed in the following pages.[2]

As will be discussed later, Crandall High embodies the kind of opposition to the reform movement that gained momentum state-wide. According to the District officials, Crandall's Principal came to oppose the outcomes incorporated in the Act, and, in addition, Crandall teachers became overwhelmed and disenchanted by the pace and the workload of the reform efforts they had undertaken. In addition, many Crandall teachers felt that they had not been consulted when

---

[1] From the *Legislative Bulletin.* (March 23, 1995). Oregon Education Association & Oregon Association of Classified Employees.

[2] The CIM outcomes are to be replaced with subject area standards. The assessment system is to consist of valid reliable assessments that are multiple-choice and open-ended, but keyed to subject area standards. To receive a CIM, students must achieve ". . . a high degree of mastery in mathematics, science, history, geography, economics, civics, English, a second language, and the arts." (p. 1, *Oregon's school improvement: Moving forward.* Oregon Department of Education. July 26, 1995.)

District A chose to develop tasks for one of the CIM outcomes. Crandall teachers, thus, refused to develop and field-test assessment tasks as part of the pilot project.

EXHIBIT II

Certificate of Initial Mastery Outcomes[3]

A student should be able to apply foundation skills and core applications for living as follows:

- **Foundation Skills**

  Think. Think critically, creatively and reflectively in making decisions and solving problems.

  Self-Direct Learning. Direct his or her own learning, including planning and carrying out complex projects.

  Communicate. Communicate through reading, writing, speaking, and listening, and through an integrated use of visual forms such as symbols and graphic images.

  Technology. Use current technology, including computers, to process information and produce high quality products.

  Quantify. Recognize, process, and communicate quantitative relationships.

  Collaborate. Participate as a member of a team, including providing leadership for achieving goals and working well with others from diverse backgrounds.

- **Core Applications for Living**

  Deliberate on Public Issues. Deliberate on issues which arise in our representative democracy and in the world by applying perspectives in the social sciences.

  Understand Diversity. Understand human diversity and communicate in a second language, applying cultural norms.

  Interpret Human Experience. Interpret human experience through literature and the fine and performing arts.

  Apply Math and Science. Apply science and math concepts and processes, showing an understanding of how they affect our world.

  Understand Positive Health Habits. Understand positive health habits and behaviors that establish and maintain healthy interpersonal relationships.

---

[3] Taken from *Toward Implementation of the Oregon Educational Act for the 21st Century. Working designs for change.* (January 1993). Report of the Oregon State Board of Education to the Oregon Legislative Assembly.

Because of such resistance, reforms proceeded with caution. In 1993-94, only general CIM outcomes had been established. (See Exhibit III for CIM outcomes.) Thus, in 1994-95, the Oregon Department of Education (ODE) proceeded to define and to delineate curricular frameworks, which are intended to align CIM outcomes with content standards. (The curricular frameworks contain 36 content goals, 12 through 21 of which pertain to math and science. These content-goals are directly relevant to the work District A has undertaken.)

While the most intense focus during those two years was on CIM outcomes, work also progressed with respect to CAM outcomes. CAM builds upon CIM, but the focus of the outcomes were (and are) within specific endorsement areas.[4] These areas are: (1) Arts and Communications; (2) Business and Management; (3) Health Services; (4) Human Resources; (5) Industrial and Engineering Systems; and (6) Natural Resource Systems. Students not only must continue to work on the CIM Foundation Skills, but they also must demonstrate mastery in their endorsement areas. The emphasis of the CAM system is on school to work transition.[5]

The Oregon Department of Education also spearheaded several other reforms within the state's educational system. By the fall of 1995, every district is expected to have non-graded primary schools, and each school is expected to have site-based decision-making councils, called *21st Century Schools Councils.* (The state strongly recommended these measures, but they were not required by schools.) In addition, by the Fall of 1994, all districts were to use performance assessments, and, by January of 1995, every district and every school had to submit a plan to the ODE for implementing the CIM and CAM programs. According to District A's Curriculum Director, many of these timelines are likely to slip. (District A submitted its CIM/CAM implementation plan, but a cover memo urged the ODE to reconsider the reform timeline.)

Oregon's Assessment Reform

Assessment reform, a significant element of Oregon's educational reform, is being driven by the examination of *what student outcomes are essential at what levels* (including the CIM and CAM levels). In the reformed education system, students will be expected to integrate knowledge within and across traditional disciplines. In order to monitor and assess student progress with regard to these criteria, the reform plan envisions a new, more complex assessment system that will consist of several components and will focus on continual self- and teacher-evaluation.

The goal is to develop ". . . a system that consists of scoring rubrics, a pool of tasks for each outcome, a portfolio system for gathering evidence of student work, and criteria for determining whether a student's performance satisfies the requirement of CIM"[6] All students will maintain a portfolio, which will be a collection of classroom work, projects, exhibitions, group work, self-evaluations, and results of other assessments at each CIM benchmark level (currently grades 3, 5, 8, and 10). A new assessment system is envisioned, then, as an integral feature of the educational system, eventually to be used for the purpose of student certification and system accountability. The interim purposes are to inform curricular and instructional practices, as well. (As mentioned previously, however, this plan may have to be modified in light of House Bill 2991.)

---

[4] From *Certificate of Advanced Mastery, Task Force Report* (January 1993). Oregon Department of Education.

[5] *The Certificate of Advanced Mastery. Work in Progress. Planning Document.* (January 5, 1994). Oregon Department of Education.

[6] Oregon's Educational Act for the 21st Century, Phase 2 (January 1993 through January 1994) Pg. 2.

Performance Assessment Development

The state currently is making use of available expertise at the local level to develop and field-test the new assessments. In 1992, the ODE established a Student Performance Assessment Network (SPAN) consisting of six districts in various parts of the state to develop (a) a pool of tasks for each CIM outcome, (b) a portfolio system for gathering samples of student work, and (c) criteria for determining the quality and level of student performance as measured against CIM outcomes. (Six out of 275 school districts are participating in this pilot effort.)

Each participating district was allocated $70,000 per year for 2 years (1992-93 and 1993-94) to finance the development and piloting of performance assessment tasks with respect to the CIM outcomes. According to the Assistant Superintendent for Assessment and Technology, this is ". . . not a liberal amount [but] enough for them to be able to accomplish what they need to." Additional money was given in 1994-95, but all SPAN grants will end in June of 1995.

To further aid development efforts at the pilot sites, the state sponsored workshops in the summers of 1993 and 1994, and still provides on-going help and information through five consultants.

Teachers involved in the pilot project field-tested the tasks they developed in the spring of 1994, in their districts. During 1994-95, performance tasks were used for student assessment and compiled into student portfolios. Student portfolios were then scored by teams of teachers at the school level. (The initial plan also called for rescoring by another team at the state level, to evaluate reliability and validity of the new assessment system, but this latter plan did not materialize.) (The state also participated in field-testing the New Standards Project tasks in the Fall of 1994.)

For the assessment reform venture, several organizations and individuals provided help directly to the Oregon Department of Education. ODE collaborated with Northwest Regional Labs in developing the reformed system and with the New Standards Project (NSP), to generate and field-test some assessment tasks. To conceptualize the assessment system, the state also received help from NCTM, the American Association for the Advancement of Science (AAAS), Instruction Objectives Exchange (a private organization), and experts such as Joan Herman (CRESST/UCLA) and Kemp Gregory (Psychological Corporation). The state also is a partner in the New Standards Project. (However, it should be noted that NSP CIM outcomes are different from Oregon's defined outcomes.)

District Context

In 1992-93, District A volunteered to define and field-test assessment tasks pertaining to Core Application for Living, *Apply Math and Science* CIM outcome and for Foundation Skills *Communicate* and *Technology* outcomes.

Nestled in one of south-central Oregon's scenic valleys, District A serves a mostly blue-collar community contending with recessionary times. The community depends largely upon lumbering for income, but the lumber yards are half full, and the entire community is worried about its financial viability. The school district itself (serving about 4,000 students) comprises 5 elementary schools, 2 middle schools, and Crandall High School.

In the 1992-93 and 1993-94 academic years, most assessment development work in the district was conducted by the District Student Assessment Team (DSAT), which was composed of 10 volunteer or nominated teachers representing all of the district's schools. The DSAT members developed assessment tasks for the state and, within their own schools, functioned as mentors and experts in performance assessments their subject areas. In 1994-95, a new team, Certificate of Initial Mastery Implementation Team (CIMIT), coordinated all reform activities, and a SPAN sub-committee worked on developing the CIM system. Another team, Science and Math (SCAMA) worked on establishing CIM math and science content for grades K thru 8.

By the end of 1993-94, the district plan was to develop 2 performance tasks at each benchmark level (grades 3, 5, 8, and 10) and, as part of the pilot program, to use the state-developed scoring rubric (described later) to (a) assess student performance, (b) collect the results in students' portfolios, and (c) hold an exhibition of student work toward the end of the year. (The rubric was developed by a state-sponsored committee during the Summer of 1993, and revised twice during 1993-94.) By the Spring of 1994, the DSAT team, using state grant money, had developed 2 performance tasks at each benchmark level. For their efforts, each DSAT member could choose to be paid $500 or to receive 5 days off. (Most chose to be paid. However, some who chose substitutes found that the substitutes could not teach within the applied problem-solving and cooperative learning frameworks teachers were trying to use in their classrooms.)

By 1994-95, the district pilot effort had expanded to include a total of 150 teachers and all students at the first three benchmark levels. In addition, most of the other K-7 levels participated in performance assessments and created portfolios, mostly with applied math and science tasks. By December 1994, teachers had designed 19 *applied math and science* tasks at the first benchmark level (grade 3), 9 at the second benchmark level (grade 5), 18 at the third benchmark level (grade 8), and 3 at the fourth benchmark level (grade 10).[7] The district hopes to involve all students in compiling a CIM portfolio based on CIM outcomes, and in holding culminating exhibitions of student work. Policies regarding how these hopes will be realized are yet to be determined.

Financial, Development, and Information Support

In 1993-94, the district received $35,000 SPAN grant for developing the K-12 CIM tasks for the Core Applications for Living, *Applied Math and Science* system and for the Foundation Skills *Communicate* and *Technology* outcomes. In 1994-95, this grant was reduced to $20,289, as Crandall High had withdrawn from SPAN; however, the district received a separate grant of $38,500, to be used between January 1, 1994 and June 30, 1995. District A also received $56,000 for developing the *Industrial and Engineering System* for Certificate of Advanced Mastery (CAM) in 1993-95, to be used at Crandall High. In addition to the assessment grants, the district received $14,000 for developing the K-8 CIM science and math content areas.

---

[7] From *Certificate of Initial Mastery Implementation Plan*. Central Point School District #6.

From the perspective of the state Education department, the grant money to District A is for a pilot program, the benefits of which must extend beyond the district to the entire state. Hence, one of the major tasks of the District is to disseminate the findings of the pilot program to other districts.

In 1993-94 and 1994-95, the DSAT members used their assessment tasks and the associated scoring rubrics to provide training to other teachers within the district. This exercise involved scoring student papers, choosing tasks to share with other districts, and identifying student anchor papers. The state math coordinator (a mathematics teacher) will then send score task samples (*good*, *medium*, and *bad*) to other districts. In 1994-95, the district provided six early release days to teachers for CIMIT meetings, and in April, 1995, held a *Portfolio day.*

In 1993-94, in part due to its public relations efforts, District A faced no organized community opposition to its reform activities. It was a ". . . conservative district, but one happy with the changes," according to Crandall's Assistant Principal.

This "happiness" with change seemed to be the result of a number of strategies used by district officials to foster community understanding and cooperation. The district officials:

- Utilized a *peer-support approach* to encourage the use of performance assessments;
- "Downplayed" the CIM outcomes (because those outcomes might become a bone of contention in a community that equates outcomes-based education with a "dummied-down" curriculum);
- Encouraged students to share their experiences and assessment results; and
- Provided support and encouragement to district teachers to become *trailblazers* in what is supposed to become a state system.

For example, as part of its public relations work, in February of 1993, DSAT held a district-wide voluntary workshop for its teaching staff. The DSAT members described the performance assessment tasks, rubrics, and portfolios, and attended student-led conferences on exhibitions. Students were also brought in as "co-teachers" for the workshop. According to district officials, the workshop was a success.

By 1994-95, however, Crandall had withdrawn from the SPAN grant, as, according to the District officials, Crandall's Principal and many teachers had become opposed to the CIM outcomes. The district, however, continued to use the same support mechanisms throughout 1994-95. It currently is informing its parent group primarily through organizing student-led conferences; and, reportedly, parents were enthusiastic at these conferences and provided positive feedback to the students.

Other Tests

Despite the move toward performance-based assessments, the district continues to use the multiple-choice, norm-referenced California Achievement Test (CAT), and all districts in the state must administer the Oregon State Assessments (some multiple-choice and some open ended) in the areas of reading, writing, and mathematics at grades 3, 5, 8, and 11. (Special education children are exempt from taking the state assessments. According to the District Superintendent, special education students constitute about 11 percent of the population.) The district uses the CAT for Chapter I reporting purposes and for selecting students for its Talented and Gifted Program. District officials hope to discontinue the use of the CAT sometime in the future.

Impact on District Officials and State-District Relations

District A officials believe that the primary purpose of the overall reform effort is to raise educational standards. Hence, they have enthusiastically placed their support behind the reforms. They see CIM and CAM as essentially improving teaching practices, enhancing student performance, and producing a more knowledgeable workforce.

Ultimately, officials hope to replace Carnegie units, grades, grade-levels, report-cards, and high-school diplomas with bench-mark levels and student-led conferences. Consequently, they are placing emphasis on (a) team and interdisciplinary teaching, (b) "schools-within-a-school" programs, and (c) mixed-age grouping at the elementary school level. These officials also desire to support the use of performance assessments regardless of CIM and other reforms, because they believe that such assessments result in "better teaching and better learning."

In 1993-94, because of the partnership in developing and effecting assessment reform and state financial support, the district's relationship with the state improved considerably. At the same time, individuals at the district level felt less dictated to from the state and believed they had come to have direct input into the reform process. During 1994-95, sentiments again shifted — this time toward feeling frustrated with the swift pace of reform.

The district has not yet evaluated reform processes and outcomes, but the superintendent is writing an evaluation plan. The impact of reforms in 1994-95 on district teachers and administrators is discussed in more detail in Developments in and Impact on District in 1994-95, which follows the next section below.

School Context in 1993-94

In 1993-94, Crandall High School served approximately 1250 students in grades 9 through 12, most of whom were white. About 35 percent of Crandall graduates go on to college in southern Oregon. In the past two years, according to teachers, the drop-out rate had declined, student self-esteem had improved, and students appeared to be more directed and focused in their approach to learning.

Performance Assessment Development

Fifteen people taught math and science at Crandall, two of whom (both DSAT members) were developing tasks for the *Applied Math and Science* CIM outcomes. The two participant teachers were full of ideas and had developed innovative performance tasks to use for student assessment.

Performance Tasks

Tasks developed by Crandall teachers consisted of extended projects students were required to complete both inside and outside of the classroom. For example, a one-week project required students to plan *a residential zone with a view to maximizing profits for the real-estate developer* (see Appendix A). Students could not violate certain building codes, therefore, they had to work with constraints such as residential plot sizes and number and types of drives each residence could have. In addition to the subdivision plan, students had to produce a cost-sheet showing profits.

Another nine-week project centered on the use of math and social science to study the area's watershed. Among other considerations, students had to test the *ph* levels of the water and record the area's rainfall level.

Math and Science Scoring Rubrics

A scoring rubric with score-points ranging from one to six — unacceptable to excellent — is applied to all projects utilizing principles of mathematics and science (such as the ones described above) — and is used to judge the quality of student performance. The rubric a pivotal piece in the assessment system, as it specifies the dimensions and levels of student performance in the domains of mathematics and science. This rubric was developed and revised by a state-sponsored committee composed of teachers and other professionals. In 1993-94, the rubric specified the *conceptual understanding, processes and strategies, interpreting results, communication,* and *use of technology* aspects of the *Apply Math and Science* outcome. In the revised 1994-95 scoring rubric, the last dimension, *use of technology*, was dropped, and the language defining each criteria and score-point was simplified.

This new rubric also simplified the breaks between score-points; the earlier rubric had grouped together scores 1 and 2, 3 and 4, and 5 and 6, making it difficult to judge the differences between the grouped scores. The new rubric is presented in Appendix B. (The term "scoring rubric" has been replaced with "scoring guide," as the original term upset many people. District A officials facetiously referred to the new scoring guide as "roaring scubrics.")

The scoring rubric is a generic one, equally applicable to mathematics and science tasks at all benchmark levels. It requires teacher judgement in determining what in the student's work demonstrates the presence of the defined domains of thinking, and at what level.

In an evaluation to be conducted at the end of 1994-95, the state and the district hope to address technical issues, such as *reliability*, *validity*, *skills coverage*, and *time to completion* of these assessments.

Other Assessments

At Crandall, diffusion of design and the use of performance assessments from math and science into other subjects occurred through the informal sharing of ideas among teachers. Some teachers used the example of the *Apply Math and Science* rubric to develop and experiment with innovative assessments and rubrics for their students in a variety of subjects. In addition, they also started requiring their students to compile their completed assignments into portfolios.

As an example of such "borrowing," *Tools of Our Time*, an applied technology class, utilized several performance assessment measures. This class is described in Exhibit III. Teachers also used a variety of non-academic scoring rubrics, such as the *Work Habits* rubric, to evaluate student attendance, as well as quality of work, dependability, knowledge of the job, quantity of work, team work, and initiative. (See Appendix C for a sample rubric.). Some teachers also began to require students to assess themselves and assess one another.

EXHIBIT III

Tools of Our Time

For *Tools of Our Time*, students work in teams of two or three to complete technology-based assignments in various subject areas and keep portfolios of their completed assignments. They work in a technology lab that contains an optical scanner, photography equipment, desk-top publishing equipment, a CD ROM, video clips, and computer-aided drafting tools. To illustrate their point, they may electronically clip art from a CD containing pictures of the Egyptian pyramids to accompany an essay, or they may simply use text, but project this text onto a large screen to show their work. For example, for a student demonstration of a *Tools of our Time* project in social studies, a student used an overhead projector and a computer to display his story, *What it Felt Like to be All Alone for Three Days After a Plane Crash*, to an audience in a lecture room. The story wove in concepts and terms from economics — capital, land, labor, barter, productive, resources, value, and so on. A question and answer session between the student and his teacher followed the demonstration.

Training and Resource Support

According to Crandall's Assistant Principal, various types of support are essential for change, and she attempted to provide them all. She believed, for instance, that providing moral support for staff is critical, if the new educational system is to be accepted and, ultimately, be successful within the classroom. She said that the ". . . ultimate difference is the teacher in the classroom — with a poor teacher, you can't have a fine product." She sees, then, that a fundamental change in teachers' roles is necessary for reform to work. A new technology, by itself, will not succeed. To help teachers change their roles, she observed that "Innovators need to know that if it doesn't work, nothing is going to be held against them." She maintained, too, that practical support, such as *regular time* for developing the system, is necessary. Finally, she contended that monetary support is indispensable, and innovative programs that cut across budget lines are necessary, all of which require new resources and creative reallocations of the existing funds.

In addition to the continuous moral support from the administrative staff, Crandall teachers received five half-day in-service sessions during the year. Most of the in-service sessions, on topics such as cooperative learning, were conducted with an eye to 1997—to meet the time-line for implementing CIM.

DSAT meetings also were an important vehicle for informing and training teachers about the CIM system and associated assessment reforms (all DSAT meetings were open to all teachers). For example, some tasks were field-tested during the early spring of 1994, and the DSAT members came together, in mid-Spring, to discuss the applicability of the *Apply Math and Science* scoring rubric. At the scoring session, teachers pored over a number of completed student performance tasks, and discussed *what constitutes the distinction between a score of "0" and a score of "1" on the scoring rubric* (with the discussion extending to the other score points as well). For example, one teacher said, "We have to know the difference between 5 and 4 for CIM, because if we cannot distinguish between the two, we end up in court." The teachers also discussed the importance of assigning tasks that are truly challenging. Teachers received $100 per day for DSAT training, which they typically used as remuneration for substitute teachers.

DSAT members provided in-house support and training to teachers on an informal basis. Teachers, however, received no extra time for planning assessments. Out of an 8-period day, they typically had two 45-minute preparation periods per day (for planning lessons and assessments, and for taking care of other functions). Several teachers said they valued informal support and would like to have had greater amounts of such assistance. Teachers would like to have made contact with ". . . master teachers for sharing ideas" and for "effective models of instruction and assessment." Staff at Crandall High also voiced a desire for more "research information."

Although the bulk of professional development and support activities, as well as ideas and energy, emanated from within the school, some direct state support also was available (along with help from professional organizations). Crandall High had several grants — a 2020 grant (for staff development and school restructuring), a SPAN grant, and a Developmental Site grant. Although most grants were not funneled into developing assessments, they were considered important — albeit indirectly — to assessment, as the entire educational system is being overhauled. In the summer of 1993, a team of seven teachers went to a state-sponsored training session concerning the use of rubrics for assessment. Crandall High also received some help in refining assessment tasks and rubrics from the Northwest Regional Labs (through its 2020 grant).

Design of technology-based projects in *Apply Math and Science* (and in other subjects) was an ideal assignment for Crandall teachers and students, as Crandall High's Technology Lab served as an important resource for the development of the *Technology* aspect of the *Apply Math and Science* program. The technology program was established two years ago with district and state funds. The two teachers developing the assessment tasks drew not only upon their own expertise, but upon the expertise of the technology lab. They reached out to other people as well; and, for practical ideas and information, the two teachers frequently visited projects, such as the *Delta Technology Project* at Delta Middle School in Grand Junction, Colorado.

Interaction with Other Crandall Reforms

The drive for assessment reform was being fueled by other reforms that were already underway at Crandall. Crandall, for instance, proudly established its seven Schools Within Schools (SWS), with the twin goal of improving student attitudes toward learning and encouraging teachers to employ more creative teaching techniques. By 1993-94, Applied Studies (i.e., the Talented and Gifted Program) and Crandall Alternative Program (i.e., the At-risk Program) were in their third year of operation, and the School of Social Service, the School of Business, and Community Based Education were in their second year of operation. In 1993-1994, Crandall High also opened the Humanities School and School of Rogue Ecology. By this time, about 23 percent of the students attended SWSs, and 77 percent attend regular school.

Crandall teachers' faith in SWS was supported by students' GPA and attendance data. A comparison of the SWS students' 1991-92 GPA and attendance data with their 1992-93 GPA and attendance data showed that students had improved GPAs and attendance records.[8] As of this time, however, freshmen were not eligible for SWSs, and there were no concrete plans at Crandall High for involving the entire school in SWSs.

Because teachers were doing integrated work in the SWSs, integrated project-based assignments and the use of performance assessments seemed to them a natural extension of what they already did. At all these "mini-schools," students concentrated on certain integrated subject areas for the first four periods in the day and took traditional subjects in the other SWS for the last three periods.

In keeping with the *Educational Act for the 21st Century*, Crandall also established a site-based council which is charged with (a) developing the curriculum, (b) professional development, and (c) implementing the 21st century act. It is unclear, however, how much power it had to make budgetary decisions. The council is composed of teachers, administrators, parents, and students.

Impact of Assessment Reform

As is the case during any process of change, the impact of reforms on the individuals involved was an amalgamation of the good and the not-so-good.

Impact on Teachers

The thrust for outcomes-based education reform simultaneously engendered enthusiasm and confusion among teachers at Crandall High. Most teachers and administrators, however, were delighted that the CIM and CAM tasks and standards were being defined at the local level and had not been handed down as dictates from above.

Although the proposed CIM outcomes were forcing most teachers here to take a closer look at integrative work and performance assessments, only some — about 19 out of 75 — were using integrated performance assessments as part of their daily routine, and mostly in the school within school classes. According to one teacher, most of their colleagues did not ". . . buy into the [reform] concept." For most, CIM portfolios were seen as simply too much work. The CIM outlines were believed to be too general, and teachers felt they would be "sacrificing content to teach process skills."

---

[8] From *21st Century Schools Waiver Annual Report, 1993.* Crandall High School

On the other hand, the 19 teachers who were involved had pioneered their own assessment tasks, fashioned some of their own scoring rubrics, and were using the assessments to gauge student work and to inform their own instructional strategies. Some of them also based student grades on the assessments, although report cards were presented in the traditional letter-grade format.

Anecdotal evidence suggested that teachers who were using performance assessments were being energized into thinking about the multidimensionality of knowledge, its expression, and its assessment. They were assigning "project work" and thinking through the types and levels of competencies each project should capture. These teachers believed that subject-integration offered students a better chance at internalizing what they were learning and at developing higher-order thinking skills.

At the same time, these same teachers held divergent opinions on the usefulness of the rubrics. Some considered them to be, indeed, useful on some accounts, but others also worried about "lowering standards," for they did not know how to translate the criteria at each score-point into what could be concretely expected from student work.

One teacher said, "The purpose of the rubric is to help students get feedback." Another said, "The instructional purpose of the rubric is to be able to tell the student where he or she is, what his or her progress is."

Teachers also harbored several concerns about the CIM and the associated assessment system. They were unclear about how CIM and CAM would work in tandem and about the consequences, especially for college admission, for students who do not attain the CIM and CAM levels. Many teachers were confused and concerned about the perceived lack of connection between the CIM and CAM outcomes and the traditional content areas.

Teachers were uncertain, too, about the work expectations the new system is to establish. The debate at Crandall High centered around where students and teachers fit into the state-projected reformed environment, and teachers were concerned about the substantial time demanded by project-based work and its assessment. Ultimately, they were skeptical about the system, and the question on their minds was, "Is it really better"?

The Assistant Principal was of the opinion that the assessment strategy will be sustainable only if it is infused with substantial content standards; and she added that if the curriculum placed what the public teachers believed to be undue emphasis on "how to think" and not enough on "factual" learning in the traditional disciplinary subject areas, there would be resistance to the process in the form of a "backlash."

Indeed, these concerns portended Crandall's withdrawal from the SPAN grant, and therefore, from its CIM development and piloting responsibilities. At the end of the 1993-94 academic year, Crandall's site-based committee voted down Crandall's membership. (The catalyst might well have been teachers' anger at the fact that they had not had any involvement in changing the *applied math and science* rubric.) According to the district officials, the Principal had developed strong reservations about the entire reform package, and mathematics teachers were unwilling to do "extra things," such as developing, piloting, and scoring performance based mathematics tasks. Individual teachers, nonetheless, continue to develop and score performance tasks, and work continues on the CAM strands, although its progress is slower than that of the CIM system.

Impact on Curriculum and Instruction

Although teachers were using varied instructional strategies, they did not attribute their use of such strategies to the use of performance assessments *per se*. According to teachers, many of them were employing cooperative learning and team-teaching pedagogical strategies, regardless of performance assessments. However, teachers felt that cooperative learning was more successful with project-based work than with other types of knowledge transfer. The important point was that assessment is not viewed as being "instrumental" in changing teaching strategies. "Our assessment didn't drive anything—everything drove assessment. Purpose is not assessment; purpose is to teach," said one teacher.

Impact on Students

The two student interviewees included in this study had similar responses. One said that she liked project work because she could integrate English and social studies and use computers and other technology to produce her assignments. She described her project-work as "self-motivated learning." The other student was similarly enthusiastic about learning through technology and integrating subject areas into projects. Neither, however, seemed to have paid particular attention to the grading criteria.

Teachers felt that performance assessments had a number of positive effects on their students. They observed that student participation in self- and peer-assessment compelled students to accept more responsibility for, and take more ownership of, their grades. One teacher said that her students were responding with "enthusiasm and motivation" — many felt that projects and regular assessments had resulted in higher homework completion and attendance rates. Teachers who were requiring more integrated work and performance assessments believed their students were internalizing knowledge more than they would have in a regular, traditional subject classroom. One teacher said, ". . . [traditional] tests show what kids can regurgitate. . . I'm sold on alternative assessments."

However, most teachers also were of the opinion that bright students who were not performing well on traditional tests had tended do well on performance assessments; but they said that the ". . . top 5 percent [of students who do well] stick with the traditional system [not SWS] and go on to college." Most assessment reforms were occurring primarily within Crandall's Schools Within Schools.

Impact on Parents

The district and schools made some effort to engage and inform the public in the reform process. During 1993-94, the CIM and CAM committees at Crandall High included two parents, each. Community involvement, however had been "minimal," said one teacher.

Indeed, the parent included in this study was unaware of the state's reform initiative. However, she did possess knowledge of her child's current instructional and assessment arrangement and viewed it in a positive light. She felt that her child was ". . . excited about learning and gaining knowledge." Teachers using performance assessment likewise intimated that families were more engaged in their children's schoolwork because of assignments requiring family support and input, such as the one called "Family History," for which students had to interview their family members.

Crandall's Withdrawal from Pilot Efforts

In 1994-95, Crandall withdrew from the CIM Task development program. According to district officials, Crandall's Principal stopped supporting the reform activities. In addition, Crandall teachers voted to withdraw from the pilot, as they were feeling overwhelmed with work. Although all 15 of Crandall's math and science teachers were required to develop performance tasks, field-test the tasks on their students, and share the findings with their colleagues, only two did so. These two teachers taught in one of the school-within-schools. According to one district official, the reasons behind Crandall teachers' decision to withdraw might be summarized by the statement, "we aren't model teachers and we don't have model programs."

Developments in and Impact on District A in 1994-95

This section focuses on the developments in, and impact of, the new assessment system at District A during 1994-95. During that year, according to one district official, "the going was a little rough." As a result of the reaction at Crandall, District A, thus, focused on building the CIM outcomes on traditional content and basic skills. The sections below detail some of the issues experienced by District A in taking the reform forward.

Performance Assessment Development

Elementary and middle school teachers at District A continued to be involved in designing, administering, and scoring performance assessments. In October of 1994, and in March of 1995, all math teachers field-tested the performance tasks they had devised with their students, and in April of 1995, they participated in a portfolio compilation and scoring session. Teachers brought to the session their students portfolios, which contained performance tasks designed to elicit the *Applied math and science* criteria and also the *quantify* and *technology* outcomes.

The district, thus, concentrated on developing and testing CIM portfolios. A Portfolio requirement list specifies the types of tasks each *Apply math and science* and *Quantify* portfolios must include. These portfolio requirement lists are presented in Appendix D. By the end of the academic year, CIMIT will prepare the final district portfolio plan.

A scoring rubric for the Foundation Skill *Quantify* outcome assessment tasks was developed by a state-level committee. The rubric incorporates the following dimensions: *Conceptual understanding, Processes and Strategies, Interpret Reasonableness, Communicate Reasoning.* As with the *Apply math and science* rubric, each dimension is to be rated along a one to six continuum, with criteria specified for each score-point. A copy of this rubric is presented in Appendix E.

During this year, teams of teachers also began to identify content standards for math and science, using the National Council of Teachers of Mathematics (NCTM) and the National Science Foundation (NSF) guidelines. The fully developed standards will be aligned with Oregon's Curricular Content Framework. In addition, teachers concentrated on infusing the developing content standards into assessment tasks. Thus, teachers now identify the content and CIM outcome areas that each performance task is intended to elicit. Although the work of defining content standards progressed relatively smoothly, mixed age-groupings made it quite difficult at the elementary school level.

In addition, the district officials began to think about the format of the future report cards. It already is piloting narrative report cards at the elementary school level, but is still using the letter grade system at the other levels. The plan, however, is not to shift from the traditional letter grades in the traditional disciplinary areas at the secondary level, unless Oregon's system of higher education changes its admission practices.

Resource and Training Support

The district continued to provide professional development opportunities during the summer and throughout the 1994-95 academic year. For example, the district held a CIM portfolios requirement session on June 18, 1994, with presentation topics such as *Scoring with Rubrics, Portfolio Management,* and *Electronic portfolios.* Responses from teachers attending the session indicated that they found the session to be quite useful and informative. In addition, building-level CIMIT teams meet regularly to score and share assessment tasks. The district also is cooperating with the neighboring districts to share assessment tasks.

For their part, the district superintendent encouraged teachers to forget about the mandates and just to think about what had worked for them and to try a few new ideas at a time.

Impact on District A Teachers

The elementary school teachers who participated in this study in 1994-95 expressed opinions not dissimilar to the ones expressed by Crandall teachers a year earlier. In fact, complexities in devising and scoring tasks using several rubrics had become even more overwhelming for them.

Most teachers began using student scoring rubrics to help students understand the performance and scoring criteria. Teachers said that the use of this self-assessment strategy was proving to be very useful with students, as it helped students gauge and set performance expectations. However, translating rubric scores of 1 to 6 into letter grades then had to be explained to students, and this proved to be a cumbersome process.

Teachers also noted the many difficulties they had experienced with tasks and scoring rubrics. Although they used the scoring rubrics to develop assessment tasks and to pilot some sent to them by the Oregon Department of Education, the tasks did not always tap into the dimensions listed on the rubrics. (Some teachers, in fact, tried to design tasks that would reflect all 11 outcome areas, which proved to be an ineffective strategy.)

In addition, several teachers also mentioned that scoring criteria were too complex and too unclear. They were not sure of whether they had to do "growth scoring," that is, score elementary school students at the lower levels of the scale and secondary students at the upper ends of the scale, or whether they had change their frame of reference based on the grade level of the child. One teacher highlighted this problem by asking the question, "Do we look at the exit [CIM] standard, or at the grade level standard?" Many of the scoring problems inhere in the fact that benchmarks for CIM levels have yet to be established, as are content frameworks. "We are all like children — give use some rules and boundaries," said one teacher. One teacher said that she was especially frustrated, because she felt like she was ". . . running a double curriculum," one for the reformed system and one for the California Achievement Test that the district still uses.

Nonetheless, the use of performance assessments and tasks developed around scoring rubrics have brought about instructional and some curricular changes. Several of the elementary teachers said that textbooks had become more like reference material for teaching content. Many had felt that science especially should not be based on books alone, and assessment tasks had nudged them towards "broadening out" the curriculum. Many said that they used performance-based assignments on a regular basis. For example, on teacher said that she had devised a "story-line teaching" method, where her students become characters in a story such as the "Oregon Trail" and "Covered Wagons" projects. She asks her students to write the story and then to judge it using a scoring rubric. She also mentioned that her students explain mathematical reasoning much better with the aid of a scoring rubric, something which they would not have done earlier. The crux of the whole system, she said, is that ". . . students become responsible for their own learning." Student and teacher devised scoring rubrics, in fact, have become a very popular instructional tool with many District A teachers.

Still, not all teachers have developed such facility with the system and are quite skeptical of it. For example, a self-proclaimed "old dinosaur" teacher tried his hand at a performance task, and asked his students to fashion a device that would send a marble down three feet in three or more seconds. Students invented different types of gadgets, including a tube, three feet long, filled with honey. The idea behind the project was to help students learn about speed, acceleration, gravity, and mass. The teacher, however, still does not like the *communicate* aspect of the scoring rubrics and wants to concentrate on teaching basic skills. His students, he said, perform extremely well on the California Achievement Tests, and, so, he is not likely to embrace the system.

This method of teaching and scoring is, nonetheless, not without potential drawbacks. According to one elementary school principal, the outcomes are "mind boggling" — teachers are still trying to write tasks that fit really well with the rubric outcome areas *and* that are tailored to content areas. Thus, a second major concern is that the extensive use of performance-based assignments not lead to a "hit and miss" curriculum. A third problem is that performance tasks and scoring is very time-consuming. Due to these concerns, therefore, many teachers do not hide their frustration about the fact that all their efforts might, in fact, be for naught in the coming years.

Impact on Parents and the Community

The eighth grade parent said that she "likes the basic premises of the system." She commended the fact that her child received more project-based work that required the use of reference materials and a computer. The student had not been exposed to that kind of work before. However, she also said that some of the CIM requirements, such as that of proficiency in a foreign language, are unrealistic.

This point also was noted by the District Superintendent. According to him the *foreign language* and *understand diversity* outcomes are quite controversial, and the School Board is divided on the issue of whether the reforms should continue. He noted that good things had happened, such as the use of performance tasks and scoring rubrics, but also that the pace of reforms had exhausted people and content standards must be specified to mollify reform opponents.

Future Plans

The future of Oregon's reform will depend largely upon the pending modifications to Oregon's *Educational Act for the 21st Century*. The CIM system will be reviewed, including the current benchmarks, and new implementation guidelines will be established in the near future. The current plan, however, is to assess students based on their portfolios, which would contain state assessments and locally-developed assessments. An important issue to be determined, thus, is how much help students can get from their parents, when they complete performance tasks at home. Students' portfolios will be evaluated by a panel comprising teachers other than their own, and district and state teams will perform spot-check on the quality of student portfolios. As the system stands now, it is quite likely that the implementation timelines will be extended, and the outcomes will be modified. In the meantime, the state and the district expect to establish a scope and sequence for the content and CIM and CAM outcomes this coming year. In addition, the state will determine guidelines for the types and numbers of tasks to be included in student portfolios. It also will detail a system of accountability.

In the meantime, District A has established its own interim timelines. By 1995-96, it will involve students at all grade levels in the CIM portfolios, and the portfolio content will expand to cover all CIM outcomes and curricular framework components.

Conclusions

It is too early to judge the extent and the impact of reforms, especially assessment reforms, at District A, although some lessons can be drawn from past and present experience. It is difficult to determine whether teachers' "naturally experimental attitudes" lead them to the perceived positive effects of the reforms, or whether the effects are present in some real sense. District A teachers, however, are relying on their classroom experiences in order to judge the content and consequential validity of the alternative assessments.

The scoring rubrics (state-defined and self-created) have become a basis for dialogue around defining concrete outcomes and establishing standards of performance. Because teachers are encouraging students to assess their own and other students' work using the scoring rubrics, they also are helping to make implicit assumptions and expectations of student work much more explicit. Nonetheless, certain issues remain at large.

First, despite the support given to the CIM and CAM systems by administrators and teachers alike, and the associated assessment reform, there is a chance that the reforms will fail to be fully instituted. In fact, the reforms may suffer set-backs, unless teachers and the public alike see linkages between prescribed outcomes and content areas, and understand the implications for students who fail to meet the CIM or CAM criteria. Other potential threats to a fully reformed system include (a) a lack of clarity about the CIM and CAM outcomes as well as about the scoring standards, (b) "teacher burnout," (c) time demands of designing, assigning, and scoring performance tasks, and (d) teachers' concern over the lack of articulation between high school assessments and college admission requirements (such as the Scholastic Aptitude Tests).

APPENDIX A

Residential Zone Extended Project

SUNRISE SUBDIVISION

Central Point Codes: Summarized from pages 208 - 215 and pages 240 - 243, sections 1616.16.020 - 16.24.080 and 17.16.040 - 17.20.070. (Complete codes available from your teacher.)

- STREETS
    - A. Width: 36 feet wide with 5 foot sidewalks on both sides, and 2 foot utilities easements on both sides.
    - B. Half streets are not acceptable (i.e.: 18 feet wide)
    - C. CUL-DE-SACs shall be as short as possible, be no more than 400 feet long, serve no more than 12 single family dwellings, and terminate in circular turn around.
- BLOCKS: Blocks shall not exceed 1200 feet in length.
- LOT SIZES:
    - A. 8000 square foot minimum size.
    - B. Irregular lots:
        - (1) 60 foot minimum width measured along the front building line (the front of the house)
        - (2) Average depth no more than 2 1/2 times the width
        - (3) Through lots are to be avoided
        - (4) Lot side lines should be perpendicular to street lines as far as possible
- R - 1 - 8 Development Requirements:

Min Lot Area (interior)	8,000
Min Lot Area (corner)	8,000
Min Lot Width (interior)	60 feet
Min Lot Width (corner)	70 feet
Min Lot Depth	N/A
Min Front Yard	20 feet
Min Side Yard (interior)	5 feet
Min Side Yard (street side)	20 feet
Min Rear Yard	15 feet

Costing Guide

- Standard lot value can be calculated at $1.88 per square foot

- Cost for street, sidewalks, utilities can be calculated at $118 per running foot along the street side. The developer receives 10 percent profit of this cost.**

- The developer receives 30 percent profit per square foot plus $4,000 additional for each lot.[*]

---

[*] Developer profits are purely arbitrary and fictional since they must include consideration of the original land cost and development costs. However, the square foot value ($1.88 & $118) of land and improvements calculations are based upon area sampling.

APPENDIX B

Apply Science and Mathematics Rubric

APPLY SCIENCE AND MATHEMATICS
A SCORING RUBRIC

	Conceptual Understanding *Demonstrates understanding of concepts*	**Processes & Strategies** *Uses effective processes and procedures/strategies*	**Interpret Results** *Analyzes and interprets information to reach reasonable results/conclusions*	**Communicate Reasoning and Results** *Communicate reasoning, processes, strategies, and results/conclusions*
6				
5	• Selects and uses concepts relevant to a situation • Creates a conceptual model or diagram to convey understanding • Connects concepts to other situations when appropriate	• Demonstrates effective use of processes and/or strategies • Displays evidence of clarity, organization, continuity and reasoning • Uses multiple strategies and/or procedures effectively when appropriate	• Includes connections and relevance in the interpretation • Includes alternative interpretations when appropriate • Shows strong evidence of reasoning in the analysis and interpretation • Verifies results and explanations	• Expresses the concepts and strategies/processes used to reach results/conclusions • Clearly communicates reasoning • Matches language, symbols, and communication forms to the information presented and to the intended audience • Clearly communicates connections and relevance
4				
3	• Inconsistently selects and uses concepts relevant to a situation • Creates a partial model or diagram which conveys limited understanding • Inconsistently connects concepts to other situations when appropriate	• Demonstrates moderately effective use of processes and/or strategies • Displays some evidence of clarity, organization, continuity and reasoning • Attempts to use more than one strategy and/or procedure, when appropriate, although minimally or not effectively	• Includes some connections and relevance in the interpretation • Includes few alternative interpretations when appropriate • Shows minimal evidence of reasoning in the analysis and interpretation • Partially verifies results and explanations	• Limits the concepts and strategies/processes used to reach partial results/conclusions • Partially communicates reasoning • Language, symbols, and communication forms adequately match the information presented and the intended audience • Communicates limited connections and relevance
2				
1	• Selects and uses concepts not relevant to a situation • May attempt conceptual model or diagram that conveys inaccurate, incomplete, or misleading understanding • Does not connect concepts to other situations when appropriate	• Demonstrates ineffective use of processes and/or strategies • Displays little or no evidence of clarity, organization, continuity or reasoning • Does not attempt to use more than one strategy and/or procedure when the situation calls for multiple strategies	• Does not attempt to include connections and/or relevance in the interpretation • Alternative interpretations are not included when appropriate • Lacks evidence of reasoning in the analysis and interpretation • There is no attempt at verifying results or explanations	• The concepts and strategies/processes do not support the results/conclusions • Does not successfully communicate reasoning • Language, symbols, and communication forms are improperly used or do not reach the intended audience • Communication of connections does not show relevance

APPENDIX C

Work Habits Rubrics

WORK HABIT RUBRICS

TO RECEIVE A (5/6)

1. Students will be seated and prepared for work as the tardy bell rings, 98 percent of the time.
2. Students will be in attendance 98 percent of the time, or make up all work missed.
3. Students will demonstrate "on task" behavior while class is in session.
4. Students will demonstrate cooperation with other students and the teacher while in classroom.
5. Students will bring appropriate material to class 98 percent of the time: This will include textbook, notebook, paper and a writing instrument.

TO RECEIVE A (3/4)

1. Students are seated and prepared for work as the tardy bell rings, 94 percent of the time.
2. Students are in attendance 90 percent of the time.
3. Students, at times, will demonstrate off-task behavior and are asked by teacher to "get busy".
4. Students demonstrate some uncooperative behavior either with the teacher or other students, and must be reprimanded by the teacher.
5. Student brings appropriate material to the class 90 percent of the time.

TO RECEIVE A (1/2)

1. Students are in their seats and prepared to work less than 80 percent of the time.
2. Students attendance drops below 85 percent
3. Students are often off-task and are oftentimes being reprimanded by the teacher.
4. Students are generally uncooperative
5. Student brings appropriate material to class 85 percent of the time.

APPENDIX D

Portfolio Requirements

Core Application for Living, Apply Math and Science

The portfolio must include:

1. Individual student results of all available statewide assessments in math and science which require application of content to real world situations.

2. Evidence of student performance on short, medium and long terms tasks, including multiple on demand and curriculum embedded ones.

3. Array of applications of which 20 percent or more represents integration of math and science, 40 percent or less represents mathematics exclusively and 40 percent or less represents science exclusively.

4. Evidence of breadth by including content and processes appropriate to the benchmark standard as indicated on the Curriculum Framework Continuum Items #12-21. Evidence from each one of the following categories must be present:

    a. number sense
    b. mathematical operations and procedures
    c. graphs, patterns, algebra and other mathematical relationships
    d. geometry
    e. measurement
    f. probability
    g. facts, concepts, principles and theories from physical science, from earth and space science and from life science (all three must be in evidence)
    h. inquiry process
    i. connections within science and between science, math or technology/engineering
    j. interrelationships between science, technology and culture

5. Evidence of depth of demonstrating performance of a complete set of ingredients from at least one of the math Content Continuum items (#s 12-17) and at least one of the science items (#s 18-21).

6. At least one piece of evidence that demonstrates standard level of performance on all four scoring guide dimensions for a single task/project.

Requirements for the Foundation Skill, Quantify

The portfolio must include:

1. Individual student results of all available statewide assessments in math and science, both open-ended and multiple choice.

2. Evidence of student performance on short, medium and long terms tasks, including multiple on demand and curriculum embedded ones.

3. Evidence of scored performance on tasks including content and processes appropriate to benchmark standard as indicated on the Curriculum Framework Continuum Items #12-17. Evidence from each one of the following categories must be present:

    a. number sense
    b. mathematical operations and procedures
    c. graphs, patterns, algebra and other mathematical relationships
    d. geometry
    e. measurement
    f. probability

4. At least one piece of evidence that demonstrates standard level performance on all four scoring guide dimensions for a single task/project.

NOTES: Since evidence of work on single task may meet a number of the requirements, its is only practical that many tasks be broad in scope. However, there should be evidence of consistently being able to perform at the required level.

The requirements are the sam for all four levels. The difference is in the required level of complexity of the content knowledge and skills as determined by the Content Curriculum.

Similarly, the scoring guide is the same for all levels. Here too, the difference between levels lies in the content. This does not mean that student language or even partial scoring guides are not useful instructional tools. However, legitimate scoring should be based directly on the state scoring guide.

Kentucky Assessment Reform:
Breckenridge Middle School
May 3-4, 1994
March 9-10, 1995

KENTUCKY ASSESSMENT REFORM:
BRECKENRIDGE MIDDLE SCHOOL

Introduction

Breckenridge Middle School, part of the Berry County School District, was chosen as a site from which to study Kentucky's new performance-based assessment system, which is a part of the *Kentucky Instructional Results Information System* (KIRIS). Breckenridge serves 860 students in grades 6 through 8, 20 percent of whom are ethnic minorities, primarily African American. Berry County's population is socio-economically privileged relative to many Kentucky counties. Only 10 percent of the students at Breckenridge receive Chapter 1 services. State-wide, about 23 percent of Kentucky students are eligible for Chapter 1 services.

In 1990, just before the KIRIS system was implemented, Breckenridge experienced some major demographic and operational changes. As a result of county redistricting, the school's demographic character changed dramatically, changing from an almost exclusively white, upper income population to a racially mixed, solidly middle class population. In addition, the school changed from a junior high (grades 7-9) to a middle school (grades 6-8) and undertook a major physical renovation to support a middle school instructional configuration. In short, Breckenridge has spent the last five years adjusting to numerous changes at once — but appears to have done so relatively successfully, as the following pages will attest.

Participants

The people mentioned in Exhibit 1 were interviewed for this case study.

Observations

Observations included student preparation for and the administration of the KIRIS assessment performance events (in 1993-94) and a classroom observation of student portfolio preparation (in 1994-95).

State Context

By the late 1980s, Kentucky's education system was in crisis, with statistics showing Kentucky ranking near the bottom among states in (a) per pupil expenditures on education, (b) high school graduation rates, and (c) adult literacy. Many poor districts in Kentucky spent less than half as much as wealthier districts on each child's education; and at that time, the system showed no signs of improvement.

Assessment Development

On June 8, 1989, therefore, Kentucky's Supreme Court took steps to remedy an educational system deemed inadequate and inequitable. The court directed Kentucky's General Assembly to re-establish an education system that complied with Kentucky's Constitution. In

EXHIBIT I

Study Participants

1993-94	1994-95
• Reading, arts, and humanities consultant, Kentucky Department of Education (KDE) • District director of assessment • The principal • The school testing coordinator (who also serves as guidance counselor) • The district language arts resource teacher • Five eighth grade teachers • Four students • One fourth and seventh grade parent, and two eighth grade parents	• School board member • District director of assessment • The principal • Three eighth grade teachers (language arts and mathematics) • The mathematics department chair (7th grade teacher) • The school's union representative (6th grade science teacher) • Four students • One 6th, 7th, and 8th grade parent each

1990, the legislature responded to the court's mandate by enacting the Kentucky Education Reform Act (KERA), which adopted six broad learning goals for all Kentucky students. The goals stated that upon high school graduation, students should be able to:

- Apply basic communication and math skills in situations similar to what [sic] they will experience in life;

- Apply core concepts and principles from mathematics, science, social studies, arts and humanities, practical living studies, and vocational studies to situations similar to those experienced in life;

- Demonstrate self-sufficiency;

- Demonstrate responsible group membership;

- Think and solve problems; and

- Integrate knowledge across disciplines.

A specific provision of KERA was that the State Board for Elementary and Secondary Education develop and implement a statewide, primarily performance-based, "high stakes" assessment program to measure whether students had achieved the school goals at the expected level set by the State Board. This KIRIS assessment system also was to be used to drive curricular and instructional changes.

The KIRIS assessments are administered annually to students at selected grade levels, and all schools in Kentucky are held accountable for their students' performance on these assessments. By legislative mandate, KIRIS is used to grant economic rewards to schools that demonstrate a "threshold" level of improvement over a baseline score[1] and to deliver state assistance and sanctions[2] to schools that do not attain their threshold level.

As a way to frame 4 of the 6 goals (the first and the last two) in measurable terms, a series of 11 task forces — comprising teachers, school administrators, college professors, and representatives from the Kentucky Department of Education (KDE) — developed 75 *valued outcomes*, which spelled out what students should learn and be able to actually *do* with their knowledge and skills.

About half of the valued outcomes (renamed "academic expectations" to avoid criticism by some groups) are assessed by the new KIRIS program. Although the state is committed to using KIRIS for school accountability purposes it, as previously mentioned, hopes to influence instructional practices as well. The two "middle" goals are currently not measured, as — according to the KDE consultant — they are considered "value-laden" and, hence, too controversial to include in the assessment system.

Assessment Characteristics

The new KIRIS assessment system has three parts: (1) a battery of paper and pencil subject tests containing multiple choice and open-ended short-essay questions; (2) portfolios that present each student's best work in language arts and mathematics collected throughout the year; and (3) a small number of performance tasks that call for students to solve simulated, real-life problems, working in groups for part of the tasks. The KIRIS assessments are administered to 140,000 students in grades 4, 8, and 12 each year, across the state.[3] (All students in grades 4, 8, and 12 must take the assessments.) School districts may elect to have their students in other grades take voluntary continuous assessments, or "scrimmage" tests to prepare them for the high-stakes testing in grades 4, 8, and 12. Sixth and seventh grade students at Breckenridge, for instance, take these scrimmage tests each year.

The basic questions for all of the paper and pencil assessments and performance events are developed by Kentucky teachers. The assessments and events are, in turn, fine-tuned and field-tested by the state's testing contractor, Advanced Systems. The assessments require students to show how they can apply what they learn in real-life situations, and are intended to assess higher-order skills of critical thinking, problem-solving, and written expression.

---

[1] Baseline and threshold levels are set individually for schools and recalculated every two years. Each school must improve 18 percent of the difference between its baseline score and 100.

[2] In 1994-95, in Berry County, sanctions included requiring individual schools to prepare and adopt improvement plans and the provision of a "distinguished educator" to schools to assist them in their instructional efforts.

[3] In 1993-94, KDE was considering testing 11th graders instead of 12th graders in subsequent years, as general impressions indicated that high school seniors were unwilling to take the tests seriously. In the 1994-95 school year, students in both the 11th and 12th grades took the KIRIS test as an interim measure until the legislature meets next year to consider the issue and make a final ruling.

Because they are based upon the philosophy that all students can learn, the KIRIS assessments are currently required of all Chapter 1,[4] special education students, and students with disabilities. Special education students with Individual Education Plans (IEPs) receive test accommodations commensurate with their IEPs, such as having test questions read to them. According to the district testing coordinator, only a very small population of severely or profoundly learning disabled students (approximately 50 students in Berry County in 1993-94) are allowed to submit an alternate portfolio in lieu of the three standard test components. Below, each of those KIRIS components is delineated.

Written Test

Paper and pencil assessments, sometimes called "transitional tests," were administered in 1993-94 to students in the three "accountability" grades in four disciplines — social studies, mathematics, reading, and science. In 1994-95, a fifth subject area, incorporating "practical living", the arts, and humanities, was added to the transitional tests.[5]

In 1993-94, students addressed 28 multiple-choice questions and 7 open-ended questions per content area of the transitional tests. In 1994-95, multiple choice questions will be eliminated, except for those contained in a small research section on some forms of the reading test[6]. Students will again complete seven open-ended questions per content area.

The tests are administered over a two-week time period, so that only one subject area is assessed per day. Each subject area assessment is 90 minutes in length, with an additional 45 minutes available for any student requiring additional time. However, in practice, every student is given as much additional time as he or she needs, ". . . provided they are constructively making progress toward completion of the test. . . ."[7] For instance, some students at Breckenridge took over four hours to complete each test, according to teachers interviewed.

Assessments are composed of common questions asked of all students as well as of matrix-sampled questions that vary according to test form.[8] In 1993-94, total scores for each student were derived by adding the raw score for the multiple-choice questions to the product of four times the sum of the common item open-ended scores. (Scores on multiple-choice items were not included in the school accountability index.) In 1994-95, scores will be based solely on the open-ended questions.

---

[4] In 1993, KDE received approval from the U.S. Department of Education to use the KIRIS in Chapter 1 evaluation.

[5] In 1993-94, some questions integrated the subjects of arts and humanities and practical living/vocational studies.

[6] The research is designed to explore if the completion of multiple-choice items assist students in completing open-ended questions, or if student performance on these two types of questions is independent of one another.

[7] Kentucky Department of Education, KIRIS 1993-94 District Assessment Coordinator Implementation Guidebook, August 1993, 2-3.

[8] Of the 28 total open-ended questions in the original four subject areas, 20 are common, and 8 are matrix-sampled. The new fifth subject area will consist of four common items and 3 matrix-sampled items.

Portfolios

Under the supervision of their teachers, students spend the school year developing KIRIS portfolios that are an integral part of their classroom instruction. In 1993-94, students completed portfolios in language arts and mathematics, but only the language arts portfolios were used for "high stakes" school accountability purposes. In 1994-95, the mathematics portfolio, too, was included in the calculation of school performance. KDE has not yet determined whether KIRIS will ultimately require portfolios in other disciplines; and, according to the district assessment coordinator, a cross-disciplinary portfolio is also anticipated "in the distant future."

Eighth grade language arts portfolios. For the 1993-94 school year, the required contents of the eighth grade language arts portfolios were as follows:

- Table of Contents;
- One personal narrative;
- A written reaction or response (informative and/or persuasive) to a cultural event, public exhibit, sports event, media presentation, or to a piece of writing, current issue, math problem, or scientific phenomenon (this requirement was eliminated in 1994-95);
- One piece of writing that achieves any one or more of the following purposes: (a) predicts an outcome, (b) defends a position, (c) solves a problem, (d) analyzes or evaluates a situation, person, place, or thing, (e) explains a process or concept, (f) draws a conclusion, or (g) creates a model;
- One short story, poem, play/script, or other piece of original fiction;
- A personal selection: one additional, original piece of writing; and
- A Letter to the Reviewer discussing what he/she has learned from keeping a portfolio, which entry is the best piece and why, and from which entry the most was learned.

Any of the portfolio entries may come from subject areas other than English/language arts, but a minimum of one piece of writing must come from another area. In 1994-95, Breckenridge's curriculum committee asked all teachers in 8th grade, including those teaching electives, to ensure that students develop at least one annual contribution to their language arts portfolios in their courses. It is hoped that this practice will reduce 8th grade language arts teachers' burden and will encourage writing across the curriculum. In addition, 6th and 7th grade language arts teachers are encouraged to develop portfolios in their courses and save pieces for the 8th grade "high stakes" portfolio. Eighth grade language arts teachers indicated that as yet, however, they are receiving very few contributions from other teachers.

Mathematics portfolios. In both years of the study, the contents of the mathematics portfolio were not as strictly prescribed, but students were required to submit seven "best pieces" that ". . . represent various types of mathematics; employ a variety of mathematical tools such as calculators, computers, and manipulatives; and integrate core concepts within mathematics and the real world." In addition, students were asked to provide a "Table of Contents" and a "Letter to the Reviewer." Eighth grade mathematics teachers had developed a number of creative portfolio "prompts": for instance, students are asked to design a park that includes a budget, a map of the park's features according to scale, an evaluation of insurance bids, and so on. In another exercise, students are asked to analyze real family grocery bills and draw conclusions about the family's eating habits and the ages of family members.

Eighth grade mathematics teachers have been more successful at obtaining portfolio entries from their colleagues than have language arts teachers; department policy requires 6th and 7th grade math teachers to develop portfolios and pass three entries per student on to teachers at the next grade level.

Performance Events

Performance events are used to assess students in arts and humanities, mathematics, social studies, science, and practical living/vocational studies, but can incorporate knowledge and skills from more than one discipline at a time. Titles of performance tasks have included: "Water Pollution," "Selecting a Career," "Electoral Votes," "Heart Healthy Choices," and "Town Planning." (See Appendix A for a sample event.) An observation of a 1993-94 performance event administration follows in Exhibit II.

KIRIS Scoring

Each student taking the KIRIS receives separate scores on the transitional tests and portfolios in one of four performance categories: *novice, apprentice, proficient, or distinguished.* None of the assessments may be used to make decisions about individual students, although every student receives his or her own scores. Only an aggregate school score is provided on the performance events, however; students do not receive individual scores on performance events.

For the purpose of instructional feedback, Kentucky schools receive item-level analyses of individual students' performance in each content and skill area of the transition tests. Student attitude survey responses that are included in the tests also are reported to schools. Unfortunately, however, this feedback is received too late (typically in September) to be of much benefit to teachers in helping them to tailor their instructional approaches to meet individual student needs.

Assessment Consequences and Quality

As noted above, the state uses KIRIS test results to assess how well schools are doing and to "reward" or "punish" them accordingly. A school's accomplishment is measured by an "accountability index", which is a composite of six equally-weighted component scores. In the first KIRIS biennium (1992-94), five of the components were cognitive, including: reading, mathematics, social studies, science, and writing. The one noncognitive component is a composite of attendance, retention, dropout rate, and (for high schools) "transition to adult life." In the first

EXHIBIT II

Observation of Performance Events Administration

In 1993-94, administration of performance events at Breckenridge was observed. Students were assigned to 8 groups of 4 for the examination, which required approximately 45 minutes for the group discussion and 15 minutes for individual written responses. Each student participated in only one task.[9] The performance event was monitored by an Advanced Systems facilitator.

The tasks administered involved a short experiment or puzzle that students first investigated in groups. For instance, they were asked to test which of several instruments is more effective at separating oil from water in a simulated "oil spill," to brainstorm four kinds of jobs teenagers could fill to assist the elderly, or to discuss the areas of different geometric shapes on a *geoboard*. Students were then asked to construct individual responses that discussed the process their group followed and the reasoning behind the conclusions they drew. They also were asked to describe the experiment's application to real life, and to discuss in their responses whether and why they agreed or disagreed with their group's conclusions.

The students in each group worked well together and discussed their tasks without any disruptions. Students' responses after the assessment was that it was "easy," and that they "like it better than classes."

biennium, schools were rewarded or sanctioned based on point differentials between their biennium accountability index and the biennium point "threshold" set for each school by KDE.[10]

Schools are responsible for the scores of every accountability grade student; students who are absent (and do not take a make-up assessment during a two week window) reduce the school's overall score. In the future, dropouts also may be assigned zeros for all assessment areas, thereby lowering a school's assessment results.

In instituting the KIRIS system, the KDE abolished the old standardized, norm-referenced Kentucky Essential Skills Test (KEST). The transitional tests are referenced to state norms, and, as specified in the KERA, the assessments should be linked in some way to a test such as the National Assessment of Educational Progress (NAEP) so that educators can judge Kentucky students' achievement against national norms.[11] In its 1992-93 technical report, KDE provided national comparisons for two subject areas, and KDE has indicated that it plans to issue further comparisons of KIRIS results with NAEP results when future NAEP results become available.[12]

---

[9] This is generally the case, except in very small schools where students take two tasks to collect enough data for school-level performance reports.

[10] Thresholds were established using 1991-92 baseline scores that reflect the percentage of the school's students scoring in the *proficient* range. In order to avoid sanctions, schools were expected to achieve, over a 2-year period, an average improvement of 18 percent of the difference between their baseline score and 100. Beginning in 1992, "baseline scores" and "improvement thresholds" are recalculated each biennium.

[11] Debra Viadero (June 8, 1994) "The Little Firm That Could." Education Week, 25-27.

Although individual student scores are not publicly reported, overall school scores appear in the local media and are the subject of much public discussion. Breckenridge Middle School came very close to meeting its improvement threshold after the first year of high stakes KIRIS testing (1992-93), and in 1993-94, the school successfully exceeded its accountability threshold by 2.9 points, although this was not enough of a differential to receive a monetary reward from the state.

In Berry County, 38 percent of elementary schools, 67 percent of middle schools, and 40 percent of high schools met or exceeded their 1992-94 biennium thresholds. Nineteen of the county's 51 schools received monetary awards as a result of their strong performance. By a vote of the school staff, individual schools may choose how the financial rewards are used and distributed: some may use the rewards to provide staff bonuses, others may spend the money on staff development or school improvements for the students. In addition, the district as a whole received a monetary reward for its overall achievement, the proceeds of which will go to certified Central Office staff and to the county's alternative schools.

In the state as a whole, 90 percent of Kentucky's students scored below the proficient level in reading, mathematics, science, and social studies during the baseline year (1991-92), although scores purportedly improved in the next two years of testing.

Not surprisingly, the high stakes KIRIS system has generated tremendous public controversy and skepticism. Many in the state criticized the system as unrealistically rigorous, potentially unreliable, and certainly inequitable. For this reason, an in-depth independent evaluation of the KERA and KIRIS system was commissioned by Executive Order last year to measure its "impact on students, individual schools, school systems, and educators". The evaluation was conducted by The Evaluation Center at Western Michigan University for the Kentucky Institute for Education Research. Authors of the study made the following points:

- "The accountability index is influenced by factors beyond a school's control, but these are not taken into account when the index is interpreted. . . . Among the factors not considered are adequacy of resources, changes in the economic climate of a community, and changes in student mobility. However, the state maintains a mechanism by which a school's authorities can appeal such matters". (p. 7)

- "There is concern, but as yet limited evidence, about whether the administration of rewards and sanctions is fair to schools with large numbers of economically disadvantaged students, high turnover rates, or a very small number of students. We understand that KDE plans to provide further information on this important question in the future." (p. 7)

---

[12] The Evaluation Center, Western Michigan University for the Kentucky Institute for Education Research, An Independent Evaluation of the Kentucky Instructional Results Information System (KIRIS), January 1995 Frankfort, Kentucky:

- "The Commonwealth should investigate and report whether inner-city urban schools are being unfairly sanctioned because they have a more difficult educational task than the more stable schools."

In Berry County, the elementary school with the highest percentage of students receiving free lunches received the maximum reward possible in 1993-94. Nonetheless, the district has recognized the need to give greater support to schools with larger disadvantaged populations, and, under its "Academic Assistance Model", provides several thousands of dollars of extra annual support to schools that exhibit a combination of high student mobility, free lunch program participation, low school attendance, and low baseline KIRIS scores.

Perhaps most importantly, the independent evaluation raised the still rhetorical question of "whether the system of rewards and sanctions will help improve the quality of education in Kentucky."(p. 7) Through survey methodology, the authors found that district assessment coordinators throughout the state think that rewards and sanctions will improve education, teachers think they will not, and superintendents are divided on the matter. (pg. 7)

Nonetheless, given the necessity for schools to perform well on the KIRIS, KDE has taken great pains to guarantee the tests' reliability and validity, with somewhat mixed results to date. To meet KDE's demand for reliability, the transitional assessments and performance events are scored by temporary employees hired and trained by KDE's contractor, Advanced Systems. Advanced Systems' test development staff and content areas advisory committees write the scoring rubrics for each task, and Advanced System's staff selects student anchor papers.

With the assistance of Advanced Systems, KDE has developed an elaborate system for ensuring the reliability of its portfolio scores. Initial scoring of the portfolios is completed by classroom teachers. All teachers in the state are asked to use one of four scoring options, each of which involves some kind of blind scoring approach.[13] In 1994-95, language arts teachers at Breckenridge began to use a double-blind scoring approach to increase the reliability of their own scoring, although this approach has not yet been adopted by mathematics teachers. All teachers are also required to use holistic portfolio scoring guides developed by KDE to improve the reliability of their scoring. (See Appendix B.)

In 1992-93, a random sample of each teacher's portfolios were assigned for rescoring by other teachers in the district. Advanced Systems then identified discrepant scorers[14], and these scorers were informed of their differences and asked to attend retraining sessions. Copies of all portfolios originally scored by teachers identified as inconsistent scorers were sent to Advanced Systems and rescored by teachers who had demonstrated their skill in scoring. These rescores served as final scores for the portfolios. In 1994-95, however, this approach was abandoned because Advanced Systems found that over 70 percent of the teachers who had originally been identified as discrepant scorers turned out, in fact, to have scored their portfolios accurately.

---

[13] A blind scoring approach simply means that portfolios are scored by two or more individuals, each of whom is unaware of the others' assigned scores. For more information, see the scoring manuals prepared by the Kentucky Department of Education.

[14] by the mean of the differences and the mean of the absolute differences between the first scores and the rescores of the sampled portfolios for each teacher.

However, Advanced Systems is still utilizing another rescoring process to determine the extent to which scoring is performed accurately statewide. Each year, two samples are drawn to be rescored. The first is a stratified random sample of 200 portfolios per grade level, of which 50 are drawn from each performance level (*novice, apprentice, proficient, and distinguished*). The second is a random sample of half of the schools in each accountability grade per district, in which up to 100 portfolios per school are selected. These two samples are then rescored by a team of Kentucky teachers trained to score by KDE, using consensus portfolios as a means of monitoring consistency. A sample of Breckenridge portfolios was not rescored last year, but will be rescored in the summer of 1995.

Finally, Advanced Systems, using a team of New Hampshire scorers trained by the same experts who trained the Kentucky teachers (and the same consensus portfolios), rescore all of the portfolios from schools that meet the following "trigger" criteria:

- A school whose "Cluster Leader" (the scoring trainer for other teachers in the school) was identified as a discrepant and inaccurate scorer;
- Student portfolios reflect a substantial difference from scores on other test components;
- Student portfolio scores are substantially different from their previous year's portfolio scores; and
- The school is close to being judged as a school in crisis, and its scores on the portfolios could determine placement.

Last year, only one of Berry County's 52 schools was audited in this manner. If the rescoring process demonstrates that a school's portfolios have been significantly miscored, the rescores produced by Advanced Systems are considered final. Otherwise, the teachers' original scores stand.

Despite these elaborate precautions, the reliability and validity of the KIRIS assessments is still a hotly debated topic among educators and researchers in the state. George K. Cunningham, a professor of educational psychology and counseling at the University of Louisville, says there is ". . . 'a dissonance' between the broad education goals the state set for itself and the very specific questions that have appeared on the assessments."[15] Other educators have suggested that some of the test questions have been inappropriate for the age level of the children being tested,[16] a complaint echoed by Breckenridge teachers who felt the transitional tests were too difficult for many of their 8th graders. Authors of the independent evaluation cited above had the following to say:

> "Neither education and testing agencies nor the measurement profession has solved the many technical and operational problems with large-scale use of performance-based assessments. KDE and [Advanced Systems] might have preferred to proceed slowly when implementing the new performance assessment system. However, in the face of the legislative mandate and

---

[15] Ibid.

[16] Ibid.

> the press for reform in Kentucky, KDE and [Advanced Systems] postponed much of the needed research and development of assessment questions and implemented the legislatively mandated performance-based system at a very fast pace. KDE continues to work on the needed assessment research and development". (p. 2)

Specifically, the study authors felt that:

- The KIRIS assessments are "technically well crafted (the questions are clear and appropriate for the age group, the scoring rules are valid, and instructions are easy for students to follow)" (p. 8)

- Portfolios, while exhibiting "great instructional potential", are less reliable than other forms of assessment, because of inter-rater reliability problems (p. 7). Likewise, the performance events evidence far less reliability than the open-ended questions on the transitional tests (p. 9);

- A longitudinal approach (tracking the same group of individual students as they progress through the grades) would have been preferable to the cohort approach used (comparing each group of 4th, 8th, 11th, and 12th graders to those of previous years) (p. 8);

- The diversity of assessment approaches used is a strength of the scheme, because it enhances the validity of the results. Therefore, the authors feel that it was a mistake to eliminate the multiple-choice portions of the transitional tests (p. 8); and

- The reliability of the accountability index is problematic, despite KDE's assurances to the contrary, because of the statistical model employed.[17] (p. 9)

Evaluators summarized their concerns by noting that:

> "KDE and its collaborators have exerted herculean efforts and have accomplished much. . . . They have encountered problems, which is to be expected in so massive and fast-paced an undertaking on the cutting edge of technology. KDE is and must be in a constant state of innovation, trial and error testing, and refinement of the measurement system." (p. 2)

Resource and Staff Development Support

One of the barriers to portfolio scoring consistency seems to be the questionable adequacy of the training individual classroom teachers receive through the state's cluster training approach. The state supplies all teachers with training materials containing benchmarks and portfolio exemplars and with a "Consistency Self-Check Packet" containing examples of student work, to which committees have pre-assigned performance levels. However, each teacher receives only one day of workshop training from their school's Cluster Leader, who, in turn, is trained in a two-day session by regional coordinators and by Writing Resource Teachers who were originally trained by KDE and Advanced Systems. In 1993-94, teachers at Breckenridge complained that

---

[17] Specific concerns include "whether to treat items or students as fixed, how agreements among raters are taken into account, and whether student scores should be estimated with regression." (p. 9)

"six hours of training in August" to "grade the following April" was inadequate to ensure inter-rater reliability.

In 1993-94, both the district assessment coordinator, the district Language Arts Resource Teacher, and Breckenridge's principal verified that teachers were not receiving enough training to cope with the instructional demands of the KIRIS program or to score the test. The district assessment coordinator said that he had hosted a number of evening "KIRIS Summits" for teachers and administrators that year at the behest of KDE. Issues discussed at these "summits" included how to isolate areas of the curriculum that might be weak and how to prepare a scoring rubric. However, he was given only two weeks of "lead time" by KDE to publicize and arrange the summits; attendance therefore was low.

In 1994-95, opinion was more divided about whether the portfolio scoring training teachers received was adequate to ensure high inter-rater reliability. The district assessment coordinator noted that although substantial staff development resources were provided to schools from the state, little was targeted at scoring training. Last year, the county released two teachers from each school for one-half day to score together and compare scores, but funds were not available for that event in 1994-95. Teachers are granted 12 hours of paid in-service time to score portfolios or to receive scoring training, but teachers at Breckenridge noted that they have so many portfolios to score that all of their paid in-service time, plus much more of their free time, is spent simply scoring portfolios. Breckenridge's two language arts and one mathematics Cluster Leaders, however, felt the scoring training they had received was adequate and that they were capable of teaching other teachers at Breckenridge to score properly.

School Context

The following section examines the specific impact of the KIRIS assessments upon the major constituencies at Breckenridge Middle School.

Impact of Performance Assessment

Impact on Teachers and Administrators

At Breckenridge, every teacher included in the study felt the new system benefitted students but generated a great deal of stress and extra work. The complaint voiced most frequently in 1993-94 was the inadequate training teachers believe they had received about how to integrate the KIRIS assessments fully into classroom instruction and how to score the portfolios appropriately. One teacher expressed frustration at the fact that "I have received zero KIRIS training, yet I am held personally accountable for my students' performance."

As noted above, Breckenridge teachers were particularly concerned in 1993-94 that portfolio scoring was too subjective at the school. For instance, they felt that in 1991-92, their English portfolio scores were below the previous year's baseline, because English teachers with high standards scored the language arts portfolios, while in the first year of the KIRIS system, all teachers at Breckenridge scored portfolios, including special education and 6th grade teachers, whose standards are much lower. In 1994-95, Breckenridge teachers raised other concerns about scoring reliability; one said she didn't agree with the scores assigned to the Portfolio Exemplars by KDE, while two others noted that the holistic scoring guides were too complicated to be of much utility. Another felt that it was difficult to assign a reliable score to a portfolio comprised of individual pieces that may be of widely differing quality. Finally, one of the Cluster Leaders complained that while some teachers take a great deal of time and care in scoring their assigned portfolios, a few refuse to take the process seriously.

In 1993-94, teachers at Breckenridge also said they had received absolutely no instructions from the district or state about methods for gearing their instruction to the KIRIS assessments, other than the warning that they should develop "high levels of cognitive skills." The district did purchase kits of practice performance events (from Advanced Systems) and the teachers at Breckenridge initiated one-day performance-event training sessions for students with the district's Language Arts Resource Teacher. However, they felt this kind of training benefitted students far more than it did teachers. Teachers said then they would like to receive training in proper ways to develop portfolios, for instance, or to obtain materials that suggest cooperative learning and other activities that foster critical thinking and teamwork. (Even at present, state or district-level curricular frameworks are not available to guide the KIRIS preparation process.)

Because of what they perceived to be a lack of appropriate support, 8th grade teachers at Breckenridge were particularly resentful in 1993-94 of the accountability pressure KIRIS brings to bear upon them. They also were resentful that the burden of KIRIS accountability falls exclusively on the shoulders of 8th grade teachers — particularly those who teach mathematics and language arts — without any sort of compensatory reward. As one 8th grade mathematics teacher explained, "I feel like my name and personal reputation are attached to the school mathematics scores published in the newspaper," since there are only two 8th grade mathematics teachers at the school. She also said that even though her students' KIRIS scores had been "outstanding" the previous year, she had received no recognition from the school, district, or state for her own contribution to those scores. She did, however, believe she would have been reprimanded if her student scores had been poor. Other teachers felt compelled to have their students work and rework their portfolio pieces, until the pieces were deemed to be beyond the "novice" category. This practice seemed particularly widespread among special education teachers, as no special provisions are allowed for their students' portfolio entries.

In 1994-95, the picture of KIRIS instructional training and preparation was somewhat rosier. Teachers are using last year's performance event tasks to prepare their students for this year's events. Likewise, they feel they are learning to use cooperative learning approaches on a more regular basis in their classrooms, which serve as natural preparation for the performance events. By this time, teachers feel they also have a full arsenal of successful writing "prompts" to use in the portfolio development process, most of which they have created themselves and some of which have come from the district.

Most training dollars from the state are allocated to individual schools to use as their site-based management teams deem appropriate. This year, the county gave teachers four professional development days, and the state gave teachers two more days specifically for KIRIS training. In addition, teachers receive two flexible days of paid time away from the classroom to compensate for training in which they may have participated during the summer or weekends. Finally, KDE is providing KIRIS instruction over a Kentucky instructional television channel. As a result of all these resources, Breckenridge teachers said that they had received training this year in topics such as how to include disadvantaged students in the KIRIS system, how to generate open-ended writing prompts, how to encourage writing across the curriculum, and how to integrate portfolios into daily classroom instruction. One teacher noted, however, that new teachers at the school had to "begin from scratch" in understanding and using the KIRIS system, since performance assessments were not studied at all at the University of Kentucky's teacher education program.

The independent evaluation conducted by the Kentucky Institute for Education Research stated that, at least to date:

> "the legislative intent of integrating assessment and feedback into the instructional process at every grade level has not been achieved. Teachers need more assistance than the Department of Education has so far been able to provide to embed performance assessments into the instructional process, as was envisaged in the legislation."

The authors recommended that KDE expand training to help teachers incorporate "the performance tasks and higher quality continuous assessments into their regular classroom instructions" (p. 10).

In 1993-94, both parents and Breckenridge's principal were aware of the pressure on 8th grade teachers (and students) to produce strong KIRIS assessment scores and worried that teachers were becoming "burned out" by stress. One parent suggested that the assessments were an unfair way to judge the school's performance, since they really only examined 8th grade results. However, the principal felt that teachers were feeling more confident in and less resistant to KIRIS after the first year of assessment administration. In 1994-95, the principal was more optimistic about teacher preparedness and attitudes, noting that teachers in general felt "quite positive" about the KIRIS assessments and that they were beginning to believe in the system and to "change the way they think" about assessment and instruction.

The greatest enunciated cost of the KIRIS system to teachers and administrators, in both years, was the amount of time they must spend in addition to their regular responsibilities preparing for, administering, and scoring the assessments. Eighth grade mathematics and language arts teachers have carried the greatest share of this rather heavy burden; in 1993-94, each had to assemble approximately 150 portfolios and grade approximately 50, as well as realign their curriculum with the KIRIS tests. In 1994-95, 8th grade language arts teachers were each grading only approximately 15 portfolios (representing approximately 30 hours of work in two weeks time) since the principal had asked all language arts teachers and science and social studies teachers to take responsibility for grading portfolios. Eighth grade mathematics teachers, however, were each grading approximately 30 portfolios, (representing approximately 60 hours of work), since the only other colleagues with the requisite knowledge of mathematics to help them score were 6th and 7th grade mathematics teachers. Teachers suggested they should be receiving more time and more pay for the extra duties the KIRIS system required of them; in another county, they pointed out, teachers are paid three dollars for each portfolio they score.

Administrators must cope with the complex logistics involved in administering the test. They must receive and catalogue all test materials under secure conditions, distribute training materials to teachers, arrange assessment training sessions for students, make group and room assignments for the transitional and performance tests, copy all portfolios that are chosen for rescoring, deliver tests to home-bound students, and make decisions about alternate portfolios. Due to the complexity of the system and the number of players involved (i.e., schools, districts, KDE, and

Advanced Systems), miscommunication and coordination errors were frequent in the first years of KIRIS administration, although in 1994-95, the district assessment coordinator noted that many of these problems had been cleared up.

Nonetheless, according to the 6th grade science teacher, who is Breckenridge's representative to the Kentucky and Berry County Education Associations, teachers' unions in Kentucky are strongly behind KERA and KIRIS, because they feel that the systems have improved education and have wiped out the nepotism and corruption that were endemic to Kentucky school systems formerly. He said the unions are supportive of the system of school rewards and sanctions that KERA mandates, provided KDE can prove the reliability and validity of its scoring techniques.

Impact on Curriculum and Instruction

Teachers at Breckenridge also complained vociferously about the amount of lost curriculum time the KIRIS tests exact. They lose three days of instructional time to the transitional tests (2 days) and the performance events (1 day). In 1993-94, 8th grade students also had substitutes for 10 half days when their teachers were given release time to score tests.

Finally, in both years, the 8th grade language arts and mathematics teachers interviewed said they had dropped units from their curriculum in order to focus on elements of the KIRIS program (e.g., portfolio writing or cooperative problem-solving exercises). The language arts teachers said they had stopped teaching important units on grammar, sentence mechanics, and literature in order to do more creative writing in class. An 8th grade honors mathematics teacher said that the KIRIS portfolios require her to eliminate portions of her algebra curriculum that are critical to her honors students. Another said that she feels she loses a month of instructional time to portfolio preparation.

Finally, the independent evaluation noted that the KIRIS assessments:

> "[do] not provide teachers with timely feedback that is directly usable for improving classroom activities. While the index is not designed to provide such feedback, many of the educators with whom we communicated want more such feedback. . . ." (p. 7)

This observation was confirmed by Breckenridge teachers who pointed out that results received in September of the year following that in which students took the assessments provide them with no means of working with students to improve their identified weaknesses.

Nonetheless, in 1994-95 both the school principal and teachers themselves felt that 8th grade instructional strategies had improved as a result of the KIRIS program. One mathematics teacher noted that, "a lot of KIRIS is just what we in the math department and math community have been pushing for years." Another language arts teacher said that, as a result of KIRIS, she was learning to demonstrate to her students the real life application and utility of the principles she taught. Another language arts teacher said that although "KIRIS implementation is very difficult, the end result is worth the effort."

In particular, 8th grade teachers said they were now providing more cooperative learning exercises, using more manipulatives, and requiring much more writing of their students. They also integrate more "performance events" into their daily teaching — and they encourage participation in activities such as science fairs and math "Olympics," since these bear a resemblance to the KIRIS tests. Teachers did note, however, that Breckenridge seemed to be ahead of most other middle schools in the district in integrating performance assessments into the curriculum. Finally, by 1994-95, both mathematics and language arts teachers were becoming quite enthusiastic about the use of portfolios, indicating that they felt portfolios were an important tool for promoting independent learning and inquiry among students.

In 1993-94, opinion was divided among Breckenridge teachers about whether instructional changes had spread to 6th and 7th grade classrooms. Teachers also disagreed about whether the KIRIS system had affected instruction in mathematics and language arts classrooms only, or in other subject areas as well. In 1994-95, the principal and some teachers indicated that they felt teachers in other grades and disciplines were beginning to align their instructional methods with the KIRIS philosophy, although acceptance of responsibility for KIRIS among the full faculty would take time.

Impact on Students

There was some sense among both teachers and parents that the unlimited time dimension of the transitional assessments put an enormous strain on students, especially those who are particularly conscientious and concerned about performance. Teachers and parents pointed to the large number of students who took between four and eight hours to finish some transitional subject areas, and complained that this was too much pressure and effort for 8th graders to handle. Teachers felt that many students thought that the more they wrote, the better their score would be. Teachers, it followed, believed the problem could be solved by placing some reasonable time limits on the tests and by reducing the number of open-ended questions. Also, because the transitional tests were considered extremely challenging, there was some concern among teachers that students — particularly those who were used to doing well — would experience a damaging sense of failure. On the other hand, some parents and students felt that the transitional tests actually reduced pressure on students by posing questions that did not demand a "right" or "wrong" answer. One student said that the lack of time pressure also gave him more opportunity to think about his answers.

Portfolios, too, created a sense of pressure for some students, to the point that teachers have begun to notice that some parents are stepping in to help their children write their portfolio pieces at home. Pressure becomes particularly intense in March before portfolios are due, when students need to revise each piece for final submission. One student noted that this pressure leaves her little time for reflection to "learn from my mistakes." Teachers also noted that time pressure is worse for students who do not have computers at home and therefore must wait to use the school's limited number of computers during school hours.

Virtually every teacher at Breckenridge in both years of the study agreed that the KIRIS tests promoted the use of higher-level thinking skills to solve open-ended problems that had "real-life" application. Parents, too, applauded the challenging nature of the tests and agreed that students were forced to organize their thoughts, exercise logic, and use writing as a cross-disciplinary communication tool. Students themselves praised the creative nature of the tests, noting that they were "allowed to express [themselves]" and that "personal opinions matter."

Teachers and parents were particularly enthusiastic about the way the tests promoted intensive writing across the curriculum in 8th grade, although they were skeptical about whether increased writing was being encouraged in the 6th and 7th grades. A good sign, they said, was that the KIRIS assessments had prompted the principal and curriculum committee to eliminate typing classes and write a grant to develop a writing laboratory at the school. The Kentucky Institute for Education Research study confirmed this salutary effect of the KIRIS, noting that:

> "Students experienced more writing and group work under the reforms. Teachers, district assessment coordinators, and superintendents report almost unanimously that writing has improved, and the writing improvement was over and above what would have been expected of most school children of the same age."

Teachers felt students were definitely learning to express themselves better than ever in a variety of written genres and were more willing to revise and rewrite their work for inclusion in portfolios. Teachers said they learn more then they did before about their students through writing — and that writing exercises such as journals serve as good outlets for students' emotions. Opinion was divided, however, about whether portfolios had actually increased students' appreciation for and interest in writing. They also cautioned that although the KIRIS assessments encourage students to write more, they do not necessarily encourage students to write well, since grammar and mechanics receive minimal weight in the overall scoring.

Teachers at Breckenridge felt that students enjoyed portfolio writing and the opportunities for group work and problem-solving that the KIRIS performance events allowed them. Parents suggested that students were more invested in and concerned about their performance on the KIRIS components than they would be on a multiple-choice minimum competency exam. Interviews with students after administration of the performance events at Breckenridge confirmed their general enthusiasm for the performance events. Comments ranged from "it's fun being challenged" to "the problems are really easy if you take the time to think before you write." Teachers felt the KIRIS assessments were particularly instrumental in motivating students of average ability and that they were a better way to evaluate students, since the assessments asked students to demonstrate "what they knew rather than what they did not know."

Impact on Parents

In both years of the study, school and district administrators, as well as parents themselves, felt that parents had not received enough information about the KIRIS assessments. For instance, in 1993-94, one parent had no idea that her child was taking the performance events component on the day she was interviewed. Parents receive their students' performance level in the mail, but no explanation of the performance levels or of their child's weaknesses or strengths is provided. Some parents knew they had a right to see their child's portfolio[18], but they had to initiate a parent-teacher conference in order to do so. Due to an overall lack of information, therefore, parents do not feel the KIRIS system has improved their involvement in or understanding of their child's education. Parents interviewed in both years of the study have the following specific concerns about the KIRIS:

---

[18] Portfolios are kept at the school for five years after copies are submitted to KDE.

- They worry that since KIRIS is not comparable to other national testing systems, it does not properly prepare their children for the SAT and ACT college entrance examinations. One mother, however, understood that SAT and ACT tests also are moving in the direction of more open-ended and performance-based tasks.

- They are concerned about the consistency of educational approaches across grade levels, sensing that curricular and instructional strategies are often "radically different" in the "accountability" grades.

- They dislike the fact that assessment scores are received a year after the assessments are administered, making it virtually impossible for students to learn from their mistakes.

- They are concerned about the reliability of the KIRIS components. Specifically, they do not believe that teachers can score portfolios objectively. Further, they believe that individual students' success on the performance events is partially related to the caliber of the other students in their assigned groups. Finally, some have heard uncorroborated "horror stories" about schools in the county that pay students to perform well on the tests or give them a "day off" as a reward for taking the KIRIS assessments.

- They are also concerned about the equity of the accountability system, noting that schools with a higher percentage of disadvantaged students are likely to fare poorly, even though they agree that a system of school accountability is necessary.

- They are worried that, as a result of portfolios and KIRIS test preparation, students are missing a number of weeks of algebra and language arts curricula that they need for success in high school.

A number of parents interviewed said they understand that the KIRIS is "a better kind of test", because of its "writing and problem-solving aspects." Nonetheless, until they understand the KIRIS system fully, it is likely that some parents will, as one put it, ". . . worry that our children are being treated like guinea pigs."

Breckenridge's principal said that the school is trying to improve its inclusion of parents in the education of their children. He acknowledged that "KIRIS is a complex system to understand," and that "information parents receive can be confusing."

Future Plans

On balance, the Breckenridge educational community seemed to feel that the KIRIS assessments were improving the quality of education, at least in the 8th grade. The district assessment coordinator said that in Berry County as a whole, opposition to KIRIS had always seemed minimal, and was perhaps even less in 1994-95, since public attention was focused on controversial issues of redistricting and bussing.

However, the school board member interviewed, as well as Breckenridge's principal and union representative, noted that KERA and KIRIS continued to represent "political footballs" in Berry County and in the state of Kentucky. The school board member said that while support for

KIRIS had grown among the general population, opposition had also intensified in some quarters. In particular, certain politically conservative groups have criticized the assessments as focusing too heavily on "liberal values" and ignoring "fundamental academic skills." She blamed the "fast food mentality" among the public, which demands unrealistically quick and concrete results from KERA and KIRIS, especially given the large investment made in them. Finally, she expressed the concern that each year, fewer of the original champions of KERA remain in the Kentucky legislature.

As a result, she, like the rest of the educators interviewed, feels KERA could indeed be abolished when the state legislature convenes next year, before its efficacy and promise can be demonstrated conclusively. She hopes that the legislature will act wisely to give KDE the time needed to demonstrate the worth of KERA and KIRIS. She noted that positive public information about the systems will help assuage public concerns and that, specifically, it is critical that KDE "sells the public" on the reliability and validity of all three KIRIS assessment components as quickly as possible. The Kentucky Institute for Education Research evaluation echoed this concern, stating that, "efforts to improve the KIRIS need to be continued if it is to provide a defensible basis for high stakes decisions and if it is to contribute productively to improving classroom instruction." (p. 9)

In short, KIRIS' future, seems somewhat in question, despite its relatively promising infancy. Like the KIRIS assessments themselves, the future is "open-ended."

Conclusions

KIRIS is a state-wide performance assessment system developed in response to a judicious and political perception that, ten years ago, Kentucky was in a state of educational crisis. The purpose of KIRIS is to promote strict accountability among local schools for sustained improvement in the quality of teaching and learning. Students and teachers at Breckenridge have reacted favorably to the three KIRIS assessment formats (written assessments, portfolios, and performance events). Berry teachers report that performance-based assessments simulating real life situations seem to motivate students, while the open-ended nature of assessment questions promotes the use of higher-level thinking skills, intensive writing, and greater expressive creativity. Anecdotal evidence also suggests assessments have changed instruction in the classroom, at least in the accountability grades in which the assessments are administered.

On the other hand, the use of KIRIS for "high stakes purposes" imposes tremendous pressure on teachers in accountability grades, according to teachers, administrators, and parents. This accountability pressure may be a bit premature in that the assessments' reliability and validity are, according to study sources, still in question. The "cluster system" of *teacher portfolio scoring training* may be inadequate to ensure inter-rater reliability; and teachers also may require additional professional support to help improve their instruction techniques. In addition, the assessments themselves may need fine-tuning. According to many individuals interviewed, the assessments may be too difficult for most students, while their length may rob students of too much valuable instructional time. These are serious considerations and tensions that eventually will need to be resolved.

Finally, it appears that more information about KIRIS' characteristics and purposes should be provided to parents. Information is crucial to counteract negative perceptions regarding the assessments and the perception that students are being used as "guinea pigs" in a poorly conceived educational experiment.

APPENDIX A

Performance Event

PERFORMANCE ASSESSMENT FACILITATOR INFORMATION SHEET

TASK: SS/S1 "Town Planners"

GRADE 8
5 Students

OVERVIEW:

Students working together in a group will determine the best place for five different areas of development. Each student will be responsible for one of these areas. After 25 minutes, students will work alone on the rest of the task. This task can be done at a table or at desks that have been put together to form a table.

SET-UP/MATERIALS:

The following materials should be placed on the table:

- one response form for each student
- one pencil for each student
- one contour map
- one three-dimensional relief map
- one ruler for each student
- one set of colored magic markers

OTHER INFORMATION:

The facilitator should tell the students when the 25 minutes for the group work is completed. Please collect all material and return to kit.

PERFORMANCE ASSESSMENT STUDENT INSTRUCTION/RESPONSE FORM

TASK: SS/SI "Town Planners" **GRADE 8**

STUDENT NAME: ____________________

SCHOOL NAME: ____________________

GROUP INSTRUCTIONS:

You will have 25 minutes for steps 1-3 below.

1. Your group is a committee given the job of planning a city in the region depicted on the contour map and 3-dimensional relief map you have been given. You should discuss as a group the best locations for various facilities, activities, and land uses. These belong to five categories:

   CATEGORIES OF ACTIVITIES/LAND USES

   A. Industrial, such as factories, mining operations.
   B. Residential, such as houses and apartments.
   C. Recreational, such as parks, playgrounds, winter sports areas.
   D. Retail business, such as stores and malls
   E. Government services, such as transportation systems (e.g., major roads), water treatment, sewage treatment, fire and police departments, and government offices.

2. Determine as a group the person in your group who is to record information on your contour map. Then, as you reach agreement on locations for different activities, that person should show on the contour map the location of the activities. The different colored markers can be used to shade large areas for particular uses, and simple symbols can be used to show specific facilities. A legend or key should be produced on the mp giving the meaning of different colors or symbols. Activities in all five categories should be represented on the contour map.

   NOTE: DO NOT MARK THE 3-DIMENSIONAL RELIEF MAP IN ANY WAY!

   The person recording information on the map should insert the contour map in his or her response form at the end of the testing session.

3. For the remaining parts of this task, group members will be working alone. But first, each group member is to be assigned one of the five categories of activities/land uses listed under step 1 above. The group is to decide which member gets which category. Once this is done, the group members should move to their individual work areas and go on to step 4 on page 2.

After you finish step 3, you are to open this booklet and answer questions 4-6 on your own.

YOU WILL HAVE 20 MINUTES TO ANSWER QUESTIONS 4-6 BELOW.

4. Which category of activity or land use were you assigned? ______________________________

5. Explain your groups reasons for placing activities or facilities in your category where it did on the map. Include in your explanation what would problems may have arisen if they were placed in other locations on the map. Your response should include the discussion of any impacts, particularly those on the environment.

6. Suppose there is a major decline in industry in the city. Describe in detail the likely impacts of this on the other activities and land uses in the region

SCORING GUIDE
KIRIS Performance Event, Grade 8
1992-93
Task: SS/S1 - Town Planners

Group Activity

Question 2

SCORE	DESCRIPTION
4	Student places all categories of activities and land uses logically, and labels them accurately with key or legend (e.g.; industrial is not near residential or recreational).
3	Student places most categories of activities and land uses logically, with accurate labels.
2	Student places some categories of activities and land use logically, with accurate labels.
1	Student places one or two categories of activities and land uses logically. Student may or may not accurately label key or legend.
0	BLANK

Individual Activity

Question 4 - Do not score.

Question 5

SCORE	DESCRIPTION
4	Student identifies a problem that may have arisen if category was placed elsewhere on the map and includes impact of placement or possible placement on the environment.
3	Student merely identifies a problem without discussing environmental impact, or vice versa.
2	(Does not apply.)
1	Student's response indicates a lack of understanding of the impact of placement of categories.
0	BLANK

Question 6

SCORE	DESCRIPTION
4	Student describes possible impacts on activities and land use. Response is detailed and may include impacts on: economics, geography, social structures or cultural diversity.
3	Student describes some impacts on activities and land use. Response is less detailed than a 4 and may contain information from some of the areas of impact mentioned in 4 above.
2	Student describes impacts on activities and/or land use. Response is minimal and may include one or two of the impacts mentioned above in 4.
1	Student's response lists possible impacts on activities or land use. Impacts listed are unrealistic or minimal at best.
0	BLANK

APPENDIX B

Holistic Scoring Guide

KENTUCKY WRITING ASSESSMENT

Holistic Scoring Guide

○ NOVICE	○ APPRENTICE	○ PROFICIENT	○ DISTINGUISHED
• Limited awareness of audience and/or purpose • Minimal idea development, limited and/or unrelated details • Random and/or weak organization • Incorrect and/or ineffective sentence structure • Incorrect and/or ineffective language • Errors in spelling, punctuation, and capitalization are disproportionate to length and complexity	• Some evidence of communicating with an audience for a specific purpose; some lapses in focus • Unelaborated idea development; unelaborated and/or repetitious details • Lapses in organization and/or coherence • Simplistic and/or awkward sentence structure • Simplistic and/or imprecise language • Some errors in spelling, punctuation, and capitalization that do not interfere with communication	• Focused on a purpose; communicates with an audience; evidence of voice and/or suitable tone • Depth of idea development supported by elaborated, relevant details • Logical, coherent organization • Controlled and varied sentence structure • Acceptable, effective language • Few errors in spelling, punctuation, and capitalization relative to length and complexity	• Establishes a purpose and maintains clear focus; strong awareness of audience; evidence of distinctive voice and/or appropriate tone • Depth and complexity of ideas supported by rich, engaging, and/or pertinent details; evidence of analysis, reflection, insight • Careful and/or subtle organization • Variety in sentence structure and length enhances effect • Precise and/or rich language • Control of spelling, punctuation, and capitalization

B-1

SCORING CRITERIA

CRITERIA	OVERVIEW
PURPOSE/ AUDIENCE	The degree to which the writer • establishes and maintains a purpose • communicates with the audience • employs a suitable voice and/or tone
IDEA DEVELOPMENT/ SUPPORT	The degree to which the writer provides thoughtful, detailed support to develop main idea(s)
ORGANIZATION	The degree to which the writer demonstrates • logical order • coherence • transitions/organizational signals
SENTENCES	The degree to which the writer includes sentences that are • varied in structure and length • constructed effectively • complete and correct
LANGUAGE	The degree to which the writer exhibits correct and effective • word choice • usage
CORRECTNESS	The degree to which the writer demonstrates correct • spelling • punctuation • capitalization

INSTRUCTIONAL ANALYSIS

Examining instructional strengths can assist in improving writing and learning in your school. Student portfolios can provide evidence of instructional practices. This section of the Holistic Scoring Guide is provided to assist teachers in identifying sustained evidence of instructional practices through examination of student products. When scoring a student portfolio, scorers may identify any number of the instructional strengths listed below.

The sustained performance in this portfolio demonstrates that the student has applied instruction in the following areas:

○ Establishing focused, authentic **Purposes**

○ Writing for authentic **Audiences, situations**

○ Employing a suitable **Voice and/or Tone**

○ **Developing Ideas** relevant to the purpose

○ **Supporting** ideas with elaborated, relevant **Details**

○ **Organizing** ideas logically

○ Using effective **Transitions**

○ Constructing effective and/or correct **Sentences**

○ Using **Language** effectively and/or correctly

○ **Editing** for correctness

BEST COPY AVAILABLE

COMPLETE/INCOMPLETE PORTFOLIOS

A portfolio is incomplete if any of the following apply:

- Table of Contents does not contain required information
- Table of Contents does not note study area information (including the letter to the Reviewer)
- There are fewer than 7 different entries, including Table of Contents and the Letter to the Reviewer
- One or more entries are plagiarized (must be proven)
- One or more entries are different than those listed in the Table of Contents
- One or more entries are written in a language other than English
- One or more entries demonstrate only computational skills, or consist of only diagrams or drawings
- Portfolio contains a group entry
- Entries are out of order without clear descriptors on the Table of Contents

A portfolio is complete and will be scored according to how well it fulfills the criteria of the Holistic Scoring Guide if **one or more entries are:**

- out of order with clear descriptors on the Table of Contents
- questionable concerning fulfillment of the purpose for which it is intended
- questionable concerning plagiarism, but the plagiarism cannot be proven

KENTUCKY MATHEMATICS PORTFOLIO
HOLISTIC SCORING GUIDE

An individual portfolio is likely to be characterized by some, but not all, of the descriptors for a particular level. Therefore, the overall score should be the level at which the appropriate descriptors for a portfolio are clustered.

1994-95

	○ NOVICE	○ APPRENTICE	○ PROFICIENT	○ DISTINGUISHED
PROBLEM SOLVING Understanding/Strategies	• Indicates a basic understanding of problems and uses strategies	• Indicates an understanding of problems and selects appropriate strategies	• Indicates a broad understanding of problems with alternate strategies	• Indicates a comprehensive understanding of problems with efficient, sophisticated strategies
PROBLEM SOLVING Execution/Extensions	• Implements strategies with minor mathematical errors in the solution without observations or extensions	• Accurately implements strategies with solutions, with limited observations or extensions	• Accurately and efficiently implements and analyzes strategies with correct solutions, with extensions	• Accurately and efficiently implements and evaluates sophisticated strategies with correct solutions and includes analysis, justifications, and extensions
REASONING	• Uses mathematical reasoning	• Uses appropriate mathematical reasoning	• Uses perceptive mathematical reasoning	• Uses perceptive, creative, and complex mathematical reasoning
MATHEMATICAL COMMUNICATION Language	• Uses appropriate mathematical language some of the time	• Uses appropriate mathematical language	• Uses precise and appropriate mathematical language most of the time	• Uses sophisticated, precise, and appropriate mathematical language throughout
MATHEMATICAL COMMUNICATION Representations	• Uses few mathematical representations	• Uses a variety of mathematical representations accurately and appropriately	• Uses a wide variety of mathematical representations accurately and appropriately; uses multiple representations within some entries	• Uses a wide variety of mathematical representations accurately and appropriately; uses multiple representations within entries and states their connections
UNDERSTANDING/CONNECTING CORE CONCEPTS	• Indicates a basic understanding of core concepts	• Indicates an understanding of core concepts with limited connections	• Indicates a broad understanding of some core concepts with connections	• Indicates a comprehensive understanding of core concepts with connections throughout
TYPES AND TOOLS	• Includes few types; uses few tools	• Includes a variety of types; uses tools appropriately	• Includes a wide variety of types; uses a wide variety of tools appropriately	• Includes all types; uses a wide variety of tools appropriately and insightfully

PERFORMANCE DESCRIPTORS	WORKSPACE/ANNOTATIONS
PROBLEM SOLVING • Understands the features of a problem (understands the question, restates the problem in own words) • Explores (draws a diagram, constructs a model and/or chart, records data, looks for patterns) • Selects an appropriate strategy (guesses and checks, makes an exhaustive list, solves a simpler but similar problem, works backward, estimates a solution) • Solves (implements a strategy with an accurate solution) • Reviews, revises, and extends (verifies, explores, analyzes, evaluates strategies/solutions; formulates a rule)	
REASONING • Observes data, records and recognizes patterns, makes mathematical conjectures (inductive reasoning) • Validates mathematical conjectures through logical arguments or counter-examples; constructs valid arguments (deductive reasoning)	
MATHEMATICAL COMMUNICATION • Provides quality explanations and expresses concepts, ideas, and reflections clearly • Uses appropriate mathematical notation and terminology • Provides various mathematical representations (models, graphs, charts, diagrams, words, pictures, numerals, symbols, equations)	

UNDERSTANDING/CONNECTING CORE CONCEPTS

- Demonstrates an understanding of core concepts
- Recognizes, makes, or applies the connections among the mathematical core concepts to other disciplines, and to the real world

Place an X on each continuum to indicate the degree of understanding demonstrated for each core concept.

DEGREE OF UNDERSTANDING OF CORE CONCEPTS

Basic — Comprehensive with connections

NUMBER ├──┼──┼──┼──┼──┼──┼──┼──┤
MATHEMATICAL PROCEDURES ├──┼──┼──┼──┼──┼──┼──┼──┤
SPACE & DIMENSIONALITY ├──┼──┼──┼──┼──┼──┼──┼──┤
MEASUREMENT ├──┼──┼──┼──┼──┼──┼──┼──┤
CHANGE ├──┼──┼──┼──┼──┼──┼──┼──┤
MATHEMATICAL STRUCTURE ├──┼──┼──┼──┼──┼──┼──┼──┤
DATA: STATISTICS AND PROBABILITY ├──┼──┼──┼──┼──┼──┼──┼──┤

PORTFOLIO CONTENTS

- Table of Contents
- Letter to Reviewer
- 5 – 7 Best Entries

BREADTH OF ENTRIES

TYPES

○ INVESTIGATIONS/DISCOVERY
○ APPLICATIONS
○ NON-ROUTINE PROBLEMS
○ PROJECTS
○ INTERDISCIPLINARY
○ WRITING

TOOLS

○ CALCULATORS
○ COMPUTER AND OTHER TECHNOLOGY
○ MODELS/MANIPULATIVES
○ MEASUREMENT INSTRUMENTS
○ OTHERS

○ GROUP ENTRY

Grade 8

revised 9/94

BEST COPY AVAILABLE

Rite of Passage Experience:
Thoreau High School
April 12-13, 1994

THOREAU

RITE OF PASSAGE EXPERIENCE:
THOREAU HIGH SCHOOL

Introduction

This case study explores a thoroughly institutionalized and popular assessment technique developed and used at an alternative high school. Founded in 1972, Thoreau High School in Racine, Wisconsin, was the brainchild of two doctoral candidates, Jackson Parker and David Johnson, and fellow teacher Jerry Kongstvedt. These men wanted to build a high school in which students who were poorly served in traditional schools — specifically, students who were said to be "knowledge heavy but credit light" — could demonstrate their academic achievements in order to graduate. This idea of asking students to demonstrate their knowledge evolved into Thoreau's Rite of Passage Experience, or "ROPE," which requires students in their senior year to construct a portfolio of essays covering a range of topics, to write a research paper in U.S. history, and to make a series of presentations to demonstrate their proficiency in 17 subject areas.

As a leader in the area of performance assessment, Thoreau is known nationally among educators. Indeed, the school developed, implemented, and institutionalized a "performance assessment" two decades before such a phrase became common in education circles; Thoreau's principal calls ROPE "the oldest comprehensive performance assessment in the U.S." Though they tinker with ROPE regularly, Thoreau staff years ago instituted the ROPE assessment and the educational philosophy it embodies. Because of Thoreau's years of experience with performance-based assessment, the school receives numerous visitors throughout the school year. These visitors largely are educators interested in observing how they might adapt ROPE for use in their own schools and districts. Visitors also frequently include education researchers, who have written about ROPE in a number of publications.

Participants

In Exhibit I, individuals who participated in this study are identified by their roles.

EXHIBIT I

Study Participants

- Principal, Thoreau High School
- Five Thoreau Senior High School Teachers — music/ROPE, art, two English teachers, speech and dramatics/ROPE
- Three Thoreau Students — two seniors and one 1993 graduate
- One Parent (PTA leader)

Observations

Two ROPE presentations (presentations were in the areas of (1) fine arts and personal growth and (2) geography, mass media, and current events) were observed. Two classroom lessons in physics and music were observed as well.

School Context and the Rite of Passage Experience

Racine is primarily a working-class city of 85,000 people situated on the shore of Lake Michigan between Chicago and Milwaukee. Thoreau High School is one of four schools in the Racine Unified School District serving high school students.

Thoreau comprises both a middle school (grades 6 through 8) and a high school (grades 9 through 12).[1] When the school was first founded, it served only 11th and 12th graders; 9th and 10th graders were added in 1975, and the middle school grades several years later.

At the beginning of the 1993-94 school year, Thoreau's middle school enrolled 173 students: the high school started the year with 226 students. At the high school, girls outnumber boys 135 to 86 (girls outnumber boys in the middle school as well). The racial and ethnic composition of the high school during the same year was 87.7 percent white, 5.9 percent African American, 1.0 percent Native American, 0.5 percent Asian, and 5.0 percent Hispanic. Twelve students (5.4 percent) in the high school qualified for a free or reduced price lunch; 30 students schoolwide (7.5 percent) qualified. Thoreau's principal says that the school's racial, ethnic, and SES composition is representative of the Racine community as a whole.

Thoreau was founded with the goal of serving students not well served in traditional high schools. The school's founders aimed to achieve this goal by giving students control over their academic progress and responsibility for demonstrating their knowledge. ROPE is just one element of the program developed at Thoreau to reach that objective. On the whole, the school is characterized by:

- **Small class size** — Classes average 20 students, compared with 28 at Racine's other high schools.

- **Courses designed for students of all ages** — Students take courses as they judge they are ready for them, resulting in classes filled with students at all grade levels. For instance, Thoreau does not offer "10th grade English," but rather courses with titles like "The Beats" and "Hemingway." (Most courses are designed for either middle school students or high school students, not both, but some classes — most notably the school's music classes — include students ranging from 6th to 12th grade.)

- **A very wide selection of courses** — Despite its small size, Thoreau offers about 200 courses over the four-quarter school year, including geography, sociology, chess, and "the Zen of baseball." Students schedule courses for themselves through an arena scheduling system. Grades are assigned each quarter, and quarter grades (not semester grades) are recorded on students' permanent records.

---

[1] This examination of Thoreau and ROPE focuses on the high school, though information pertaining to the middle school that is relevant for understanding the school atmosphere will also be presented.

- **The Rite of Passage Experience** — Perhaps the school's most notable feature, ROPE requires students in their senior year to embark upon a year-long process of assembling a portfolio of essays, an autobiography, their employment record, book reports and reviews, and a reading list to help demonstrate their mastery of subject area material — mastery which is demonstrated further through written and oral presentations.

Together, these elements were long ago fused into a coherent educational program, reflective of the Thoreau philosophy. The wide range of course offerings support students' ability to complete ROPE successfully, though connections between ROPE requirements, course offerings, and teachers' classroom instructional and assessment strategies are not explicit. Indeed, when asked what they do to prepare students in grades 9-11 for ROPE, Thoreau teachers were unable to articulate responses. It seems probable that they could not answer the question not because they fail to prepare underclassmen for the senior-year ROPE but rather because the connections between ROPE and instruction throughout the high school were experience established years ago.

Thoreau's Student Body

Thoreau students must apply for admission to the school. Most students who gain admission to the middle school continue at Thoreau throughout their high school years; the high school also enrolls graduates from Racine's other middle schools. The middle school has had a waiting list for admission for the past four years, while the high school had a waiting list for the first time for the 1994-95 school year. Staff note that, although the school still serves students like those it originally targeted (i.e., students who did not do well at Racine's traditional high schools — students who today might be called "at risk"), the student body is becoming increasingly "middle class and mainstream."

Thoreau is associated with a magnet elementary school espousing many of the same educational philosophies embraced by Thoreau. This elementary school sends more students to Thoreau than does any other elementary school in the district.

According to Thoreau's principal, Thoreau students have higher attendance rates, higher college matriculation rates, lower dropout rates, and higher average ACT scores than students at Racine's three other high schools. Between 1981 and 1985 it was reported in the University of Wisconsin Freshman Profile that Thoreau graduates had higher freshman year college grades than any other group of students in Southeast Wisconsin. The principal also cites a study conducted by the district in the late 1980s that found Thoreau students were happier with their school than were their peers at the city's other high schools.

The Thoreau Staff

There are 31 teachers at Thoreau; most teach either high school or middle school classes, though some, such as the music teacher, teach mixed groups of students. The administrative and support staff is small: the principal is joined by a half-time assistant principal, two instructional aides, a guidance counselor, and a librarian. Thoreau's principal came to the school in 1983. In addition to his role at Thoreau, he serves as principal of another school — of about 600 students — in the district; an assistant principal oversees most of the day-to-day work at the other school.

Most Thoreau teachers have master's degrees and many years of experience. Though only one teacher remains from the original 1972 staff, many teachers joined not long after the school's founding. Teachers typically joined the school staff because they personally espoused the school's philosophy of emphasizing students' ability to demonstrate what they know and what they can do.

The Rite of Passage Experience

The Rite of Passage Experience was originally developed by Thoreau's founders and first staff members for use with just a few students. However, ROPE quickly expanded to become part and parcel of the Thoreau experience for all students as Thoreau's staff identified it to be a valuable experience for all students at the school. ROPE should perhaps be considered the defining feature of the alternative school, designed to fulfill the school's objective: To allow students opportunities to demonstrate their knowledge and achievement through written work and oral presentations, not merely through the district's and state's standardized competency tests and the accumulation of a set number of credits.

Though work on the elements of ROPE began with the opening of the school, the comprehensive assessment technique was not implemented school wide until the late 1970s, several years after the school's founding. Since that time, ROPE has evolved, but its most salient features have remained the same. In order to graduate, students in their senior year must demonstrate their mastery of subject area curricula through written papers maintained in a portfolio, a research paper, and oral presentations.

When ROPE was first formalized, graduating seniors had to demonstrate their proficiency in 23 subject areas. Students were guided through the process by the *ROPE Handbook*, compiled by teachers. At several points since its implementation, the ROPE process has been modified by Thoreau teachers. The *ROPE Handbook* was rewritten in 1984, and the number of subject areas in which students had to demonstrate proficiency was consolidated into 16. The *ROPE Handbook* was revised again in 1991, by two other Thoreau teachers, incorporating 17 subject areas.

The Rite of Passage Experience Today

ROPE requires Thoreau seniors to demonstrate their mastery in 17 subject areas they have studied during their high school careers. Some of the subject area students will have studied in discrete courses devoted to the subject (e.g., courses in biology, history, or math), while others they will have gained an understanding of through multidisciplinary course taking (e.g., a sociology course might further students' understanding in the ROPE areas of "human relations" and "multicultural awareness" or a range of courses in social studies might support students' understanding of issues associated with "mass media," "current events," and "ethics"). Students must successfully complete the ROPE process in order to graduate. As Thoreau has a waiver from district and state testing requirements, ROPE serves as a substitute for demonstrating proficiency on the four district-mandated pencil and paper tests in reading, English, government, and mathematics.

Subject Areas

Exhibit II summarizes the 17 subject areas included in ROPE and the methods through which students are called upon to demonstrate proficiency. (See Appendix A for an example of a student essay in the area of "Mass Media").

EXHIBIT II

Key Features of the Rite of Passage Experience

Subject Area	Method(s) of Demonstrating Proficiency
• English	book reports and reviews, oral presentation
• Reading	list of all books read during high school, oral presentation
• Mathematics	oral presentation
• Government	oral presentation
• Self-Expression	essay or oral presentation
• Personal Growth	oral presentation
• Ethics	essay, oral presentation
• Fine Arts	essay, oral presentation
• Mass Media	essay, oral presentation
• Current Events	essay, oral presentation
• Human Relations	essay or oral presentation
• U.S. History	research paper, oral presentation
• Science	written description of coursework completed, experiment, oral presentation
• Multicultural Awareness	essay, oral presentation
• World Geography	oral presentation
• Personal Proficiency	oral presentation
• Physical Challenge	demonstration of physical fitness, record of physical activities

The ROPE Process — First Semester

The ROPE process is roughly divided into two phases that correspond to the two semesters of the school year. During the first semester of their senior year, students take what is known as the *ROPE class*. Three ROPE teachers teach students in the essay writing process during the first quarter; during this first quarter, students write autobiographies and all of their short essays. During the second quarter, students write U.S. history research papers, and the class works on oral presentation strategies.

Students' written work — including autobiographies, topic area essays, and the U.S. history research paper — are graded by the three ROPE teachers. These teachers coordinate their work, each teacher reading essays of students not in his or her class; the grades they assign to written work are final.

Students are required to follow a strict schedule for completing written assignments (a schedule that has been in place only since the 1990-91 school year; formerly students had final, but fewer interim, deadlines). The schedule from the 1991-92 school year appears in Appendix B. (Teachers say the schedule has looked virtually the same since that school year.)

The class that supports the ROPE process has been in place for about 10 years. Prior to the introduction of the class, students paced themselves through the semester, and the school set only final, not interim, due dates. However, in the mid-1980s teachers realized that, for whatever reason, students were no longer pacing themselves as well as they once had. The ROPE class was created to remedy this problem by providing students with a more formal structure to guide them through the process.

Thoreau teachers who have been at the school for both approaches (i.e., with and without the ROPE class) say that they personally prefer the old system that required students to pace and motivate themselves. Indeed, one ROPE teacher said, "I hate teaching ROPE. I do it because I'm more 'expendable' than some other teachers." However, though teachers see the ROPE class more as a necessary evil — designed to address what teachers perceive to be a decline in student motivation and ability — than as a desirable feature of the assessment, this gripe is clearly minor.

The ROPE Process — Second Semester

During the second semester of their senior year, students give a series of oral presentations to demonstrate their proficiency in the ROPE competency areas. These presentations are made before the student's *ROPE committee*. Committee members include:

- **Two faculty members**. One of the faculty members is the student's homeroom teacher. As Thoreau students have the same homeroom teacher throughout their four years at the high school, at least one teacher on the committee is likely to know the student well. All teachers in Thoreau's high school serve on ROPE committees, working in pairs with the seniors in each of their homerooms. Most teacher pairs have between three and six seniors they work with on ROPE.

- **One Thoreau underclassman**, typically a junior, who learns about the process by serving on the committee. The underclassman is chosen by the student.

- **One or more adults from outside the school**. Students also choose this committee member; most students, according to teachers, choose a parent.

Students and their committees set the schedules for these oral presentations, the only stipulation being that all presentations must be complete by Memorial Day.

Each ROPE oral presentation typically involves both a prepared presentation by the student and a question and answer period; typically, the student presentation and the question and answer period together would last about half an hour. The teachers on the ROPE committee confirm the extent of the student's knowledge and understanding during the question period. Students say that, generally, the better the presentation, the fewer questions the teachers ask. Following the two pieces of the presentation, the committee members consult to assign the student a grade for the presentation. Exhibit III provides illustrations of one ROPE presentation.

The Scoring of ROPE Presentations

ROPE committees are fairly free to organize the work they do with their students as they see fit. Consequently, there is a good amount of variation in the ROPE process as it is experienced by Thoreau's students. For example, some ROPE committees grade oral presentations on a pass/fail basis rather than assign letter grades. Others choose to have the seniors with whom they are working give group presentations, rather than individual presentations, in some subject areas. (They cannot, however, allow group presentations in the four competency areas mandated by the school district.)

EXHIBIT III

Illustration of One ROPE Presentation

For his Fine Arts and Personal Proficiency ROPE presentation, Jeremy chose to demonstrate his knowledge of and facility in jazz music. An accomplished trumpet player who had already accepted a scholarship at a conservatory, Jeremy's choice to combine his Fine Arts and Personal Proficiency ROPE presentations was a logical one. Jeremy introduced his presentation by describing both the history of jazz and the defining characteristic of the genre. He then went on to explain the contributions of Duke Ellington to the jazz form, the musical features of the Ellington piece he had chosen to play, and the technical abilities he had to possess in order to play the piece successfully. Jeremy then played the Ellington trumpet solo (accompanied by a few peers) and asked the audience if they had any questions for him.

Witnessing Jeremy's presentation were his music teacher (who is also his homeroom teacher), the other teacher-member of his ROPE committee, and a group of younger students (the outside adult member of Jeremy's ROPE committee could not attend his presentation that day). During the presentation, members of the ROPE committee looked for evidence of Jeremy's knowledge of musical concepts and history. Because this presentation was also serving as Jeremy's Personal Proficiency demonstration, committee members asked Jeremy about why his musical achievements were important to him (i.e., his intended career as a musician).

The ROPE committee then gave Jeremy feedback on his presentation. The two teacher members of his ROPE committee had chosen to give their students only pass/fail marks on their presentations. There was no question but that Jeremy had passed this presentation. The teachers also told him how he measured up, in their opinion, against a variety of standards for musical performance: his music teacher told him that he has "an aesthetic sense beyond most other high school students."

Teachers also acknowledge — and students recognize — that they adjust their expectations for individual student performance based on what they know of a student's *aptitude*, *achievement*, and *interests*. Students known to be high achievers are expected to perform at higher levels than average or low achievers. Neither students nor teachers seem to feel that this application of individualized standards is a problem within the school community.

ROPE performance procedures and standards are not uniform across ROPE committees: each pair of teachers working together on a committee has developed its own implicit system of assessment. However, teachers and students alike seem comfortable with this fact. Although some teachers acknowledge that the application of incomparable standards across ROPE committees and across students poses a certain difficulty for the process' credibility, they do not consider the inconsistency to be a problem. Rather, teachers emphasize the value of the ROPE experience to their students — and to themselves as well.

Consequences of ROPE

All seniors must successfully complete ROPE in order to graduate from Thoreau. Students must achieve passing evaluations in at least 14 of the 17 subject areas in order to complete the ROPE requirement and graduate. In addition, 4 of those 14 passing marks must be achieved in the 4 district competency areas: mathematics, government, reading, and English. (Thoreau has a waiver releasing the school from administering the standardized exam used in Racine's other three high schools to test students' proficiency in the areas; ROPE substitutes for the standardized exam for Thoreau students.)

A few seniors each year transfer from Thoreau to one of Racine's other high schools because they find they do not want to go through the ROPE process. According to the school's principal, two students withdrew from the school for that reason in the fall of 1993; and three or four students withdrew from the school the previous year for the same reason, according to a Thoreau 1993 graduate. Of those students who stay at Thoreau through their senior year, few fail to complete the ROPE requirement.

Evaluations of ROPE

Thoreau has not undertaken any formal evaluation of the ROPE process. Teachers, however, are convinced of ROPE's value as an educational tool because of their observations of its impact on student learning and self-esteem and on their own teaching.

Resources and Training Support

ROPE is, and has been for many years, a fully institutionalized performance assessment. Its development from the time Thoreau opened its doors was an integral part of the school's practice and a reflection of its philosophy — students should be able to demonstrate what they know and what they can do. In short, ROPE is part and parcel of the Thoreau approach to education.

Financial Resources

No firm figures are available to illuminate how much it costs to implement ROPE. Thoreau gets the same amount of money per student from the district as Racine's other high schools. With that money, the school offers smaller classes, as described above, and implements ROPE with the senior class. (Thoreau does not employ any teacher aides or monitors; it is a school of teachers and students, a fact which, presumably, supports its ability to offer more and smaller classes.) Both ROPE and the school's entire approach to education are labor-intensive by design and are supported by Thoreau teachers' willingness to devote their own time to coordinating the ROPE process.

Human Resources and Staff Development

Because Thoreau experiences little staff turnover (the principal says the school has lost only six teachers over the past 10 years) and, because ROPE is so well established, Thoreau devotes few resources to formal staff development activities specifically related to ROPE. Incoming teachers know about ROPE ahead of time and are introduced informally to the process by their colleagues. Whatever they do not know ahead of time they learn as they serve on a ROPE committee during their first year at Thoreau. (New teachers are always paired with experienced teachers.)

Until a few years ago, the Thoreau staff met weekly before school. However, this practice was abandoned when a new Wisconsin mandate governing the amount of time spent in the classroom went into effect. The staff no longer had time to meet every week, and the faculty now meets as a group only once a month after school.

Until a few years ago, Thoreau teachers designed their own staff development activities, but a decline in the district's financial resources devoted to professional development has eliminated this option. The district offers teachers development activities each year on *Institute Day*. Thoreau teachers, however, express dissatisfaction with this training, suggesting that it is irrelevant to the experiences they have in their classrooms. An illustration provided by one teacher supports the point: on Institute Day in Winter 1994, Thoreau's art teacher was assigned by the computer registration to attend two seminars, one called *Farming and Elementary Education* and the other *Elementary Students with Spelling Problems*. (He chose not to contest the assignments and dutifully attended the seminars.)

Regardless of whether or not this teacher's experience is common, Thoreau staff uniformly say that they receive virtually no in-service training, and no training at all that is relevant to ROPE. However, for the most part Thoreau's experienced staff members do not seem to long for training. They are, in fact, confident that the education they provide their students, culminating in ROPE, is high quality, and they feel that the extensive informal conversing they do as a staff provides adequate "training" and room for teachers to "bounce ideas off each other."

Interaction with Other Reforms

As with ROPE, other "reforms" at Thoreau are not recent additions to the school. For example, the school's governance structure has many elements of site-based management. The faculty governance committee and the student government must both approve the rules that govern the school; for example, the school bans "boom boxes" with the approval of both faculty and students. The principal maintains veto power, but he says he has never once exercised his veto in his 10 years as principal. Interestingly, the Racine school district has recently initiated site-based management at some schools in the district, but Thoreau's staff and parents opted not to participate. In the words of one teacher, "We decided it would be foolish to take a step to do what we're already doing and open the door to interference from the district."

Thoreau staff do feel somewhat constrained from initiating some reforms they would like to see in the school. For example, several years ago the school temporarily eliminated letter grades, opting for narrative reports instead. However, this practice had to be abandoned and letter grades reinstated when the school encountered resistance from Wisconsin's university system, which did not know how to cope with the narrative evaluations. Similarly, Thoreau formerly offered more non-traditional courses, often taught by parents and even students. However, passage during the 1980's of a state law requiring that all teachers in the public schools be certified in their subject largely brought this practice to a halt.

In 1986, Thoreau joined the Coalition of Essential Schools (CES) as one of CES's original *associate* schools. According to Thoreau's principal, the school benefits from its participation in CES primarily from the publications the school receives. A few years ago, Thoreau became similarly involved in the New Standards Project. Three of Thoreau's English teachers have attended conferences on setting standards in their subject area. However, even more so than with CES, Thoreau approaches its participation in NSP with caution ("NSP stands for some things we don't," says the principal, noting in particular that most Thoreau teachers doubt the desirability of a single set of standards), and teachers say that participation in NSP is unlikely to have a significant impact on Thoreau.

Overall, Thoreau does not have plans to alter the school structure, curriculum, or assessment methods in any dramatic way any time during the next several years. After 20 years of refining the system, of making alterations that accommodate the school's student body as it evolves, Thoreau staff see no need for dramatic change. However, they expect to continue to review the work they do and to make minor adjustments as consensus suggests are desirable.

Impact of ROPE

ROPE is an integral part of Thoreau's program. As such, its impact upon students, staff, and school is pervasive. Teachers, though they quibble over the details of the process, are uniform in their enthusiasm for ROPE and its universal value. ROPE is the culmination of a student's high school education and a consuming part of a student's senior year.

Impact on Teachers

Teachers at Thoreau are positive about ROPE, feeling that it is a valuable educational experience for students and a valuable experience for themselves. As one teacher explained, "ROPE is a great thing for teachers. It fosters curriculum unification, it makes us talk about kids (even when we're mad at each other!), and it's out there as an expectation of what all kids need to be able to do. I can't imagine not doing this."

Teachers do not seem to have a problem with the fact that each ROPE committee chooses to organize its students' presentations a little differently. Thoreau was founded and joined by educational mavericks who, in the words of the principal, "took a constructivist view before it became fashionable to celebrate individuality and diversity." Quite simply, Thoreau staff seem able to accept their differences philosophically. As one teacher put it, "My first impression of the staff was that they're all a bunch of peacocks. But now I wonder if I'm becoming a peacock, too." They apparently accept disagreement because they know that they all have their students' best interest in mind at all times.

The primary, and substantial, cost to teachers of implementing ROPE is the tremendous amount of time they devote to seeing students through the experience. However, teachers interviewed said that the experience their students have with ROPE is worth the many hours beyond the regular school day teachers spend on the process.

Impact on Curriculum and Instruction

The effect ROPE has on curriculum and instruction throughout the high school is virtually impossible to examine in isolation: ROPE has been such an integral piece of the educational program Thoreau offers for so many years that its effects are inseparable from those of other aspects of the overall program.

Indeed, teachers do not identify any aspects of their teaching that they attribute to the presence of ROPE. Though they acknowledge that they might tailor instructional activities and curriculum coverage to incorporate elements beneficial to students who will go through ROPE, they assert that, in general, they focus on the subject matter, regardless of its relationship to ROPE. However, the curriculum offered by Thoreau is clearly shaped by the various elements of ROPE. For instance, ROPE has a geography component, and all students must take at least one quarter of geography if they hope to do well on ROPE's geography component. Similarly, Thoreau offers a sociology sequence, which, though not a ROPE element itself, supports the ROPE areas of *human relations*, *mass media*, and *multicultural awareness*.

One teacher identified physical education as the weakest part of Thoreau's curriculum, and correspondingly the weakest part of students' ROPE portfolios. Another teacher expressed an opinion that the school should not interfere in the area of students' lives captured by the *Personal Growth* component of ROPE.

Impact on Students

Thoreau's seniors are enthusiastic about ROPE. Several students commented that ROPE really serves as a unifying force for the senior class. Students support one another as they prepare for each presentation, and there is a sentiment that they are "all in it together."

Student performance on ROPE varies. As has been noted above, teachers typically tailor their expectations based upon their perceptions of students' ability. Consequently, ROPE portfolios and presentations that vary substantially may be given equal marks. Students are aware of this variation, but they do not seem to resent it. One student commented, "Using different standards isn't necessarily a bad thing."

Additionally, the essays students write during the first quarter of the process vary tremendously in length and scope. ROPE teachers say that most essays are 1 to 2 pages long, while autobiographies and U.S. history papers average about 10 to 12 pages each. In the half dozen or so ROPE portfolios reviewed for this case study, virtually no short essay exceeded two pages, while several were merely a few sentences long. However, some of the portfolios reviewed exceeded 100 pages.

Students feel a real sense of accomplishment upon completing ROPE. One young woman who graduated from Thoreau in 1993 told how the ROPE ceremony in May, when students receive certificates of completion for the ROPE experience (students themselves select the person or persons they would like to give them their certificate) ". . . is a much more meaningful ceremony than graduation."

In measurable terms, ROPE seems to be of significant value to students who go on to college. The recent Thoreau graduate said that she found herself well ahead of her peers in college in writing and organization skills, and continued, "The school as a whole, not just ROPE, gives us an advantage because we have freedom here. People in college get overwhelmed by choices." Another student concurred, noting that ROPE ". . . helps you discover your strengths in weaknesses in study habits."

Thoreau seniors interviewed for this case study say the ROPE process allows them to be more reflective than they could be in a traditional high school setting. Students are encouraged to think about what they know, what they do not know, what they need to know, and what they can still learn, and, in the words of one student, they "are taught to look at ourselves from different perspectives and to examine our ethical beliefs." Teachers say that students leave Thoreau with a high level of self-esteem and a feeling of accomplishment.

Teachers worry, however, about one effect of the ROPE process as it currently exists. Because students give their demonstrations in the second semester of their senior year, students who fail various ROPE components have very little opportunity for remediation. Although students may give a presentation in a subject area until they are successful, in reality, students who require more than fine-tuning of presentations are not usually able to learn what they must learn in time to pass the ROPE component in time to graduate. One student recently failed to pass one of the four competency areas and consequently had to transfer to another Racine high school to be able to graduate.

To remedy this problem, some Thoreau teachers have recommended teaching the ROPE class in the second semester of the junior year. This change would leave students with the entire senior year to demonstrate competency in the 17 ROPE areas and, consequently, more time for remediation. At least one student, however, opposed this suggestion, as she felt that ROPE is a special experience that should not be diluted by lengthening it.

Impact on Parents

Parental involvement is part of the Thoreau design of education: parent conferences are held four times each year (mandatory at the middle school level, they are optional for parents of high school students), and parents were involved in the decision, described above, not to become involved in the district's version of school-based management.

However, the parent interviewed for this case study (who is a member both of the Thoreau PTA and of the district PTA and serves on the school's grant-writing committee), says that parents are not unusually involved with the school: "Parents are no more or less involved here than they are at other schools. There are always a few parents who are very involved. But there is a substantial percentage of students here for whom the staff is literally family — the only place they get support."

Teachers express satisfaction with the parental role on the ROPE committee (most students select a parent as the outside adult member of their ROPE committee). They say that parents are frequently astonished by what their children have accomplished.

Future Plans

Thoreau teachers do not anticipate any major revisions to ROPE in upcoming years, though they will almost certainly continue to tinker with the process. Thoreau's principal, who hosts so many visitors each year and fields as many telephone calls about ROPE, says, "I would like to find a better way of sharing information about our program, to tell the story systematically. I've told the story so many times and would like to leave that as my legacy."

The Thoreau staff, says the principal, will experience a significant amount of turnover in the next five years. Current teachers say they will welcome the "new blood," and the principal is anxious to recruit several Thoreau graduates who are now teachers in the area. Just how an influx of new teachers might affect ROPE remains to be worked out, but the excitement about possibilities is palpable.

Conclusions

Thoreau High School was founded with the goal of providing students with opportunities to demonstrate their proficiency in academic areas. The Rite of Passage Experience, the school's comprehensive capstone assessment, encourages students to review and pull together the learning they have experienced during their four years of high school. ROPE encourages inquiry and promotes thoughtfulness in all domains of students' lives, academic and personal.

ROPE long has been the dominant feature of a fully integrated system of instruction and assessment espoused by Thoreau teachers. The relationships among ROPE, the school's curriculum, and teachers' instructional and assessment strategies evolved into a coherent whole years ago, making it difficult to tease out exactly how the various aspects of the Thoreau educational program interact with one another. For this reason, Thoreau provides an illustration of *what* might be achieved with an integrated performance assessment, rather than an illustration of *how* to achieve it.

APPENDIX A

One Student's ROPE Essay:

Mass Media

Mass Media

On Monday, October 4, 1993, the dead body of an American soldier was dragged through the streets of Mogadishu, Somalia. Army Chief Mike Durant was captured by Somalians when his plane was downed on Sunday, October 3, 1993. The sight of these two tragic events in Somalia drastically changed America's opinion about being involved in Somalia. Before October 3, 1993, Americans were supportive of the humanitarian effort in Somalia by about two to one. After the pictures of American casualties flashed across television screens and newspaper headlines, American's reversed their opinion about involvement in that African country. They wanted the American troops to return home by over two to one. The pictures that were brought into every American's home by mass media, effected the emotions and completely reversed the opinion of the American people. This essay will analyze mass media: its history, uses, benefits, and problems.

As defined by Webster's New World Dictionary: Media and Communications, copyright 1990, the term mass media means, "forms of communication that reach large audiences: i.e., T.V. radio, newspaper, advertisements, books, etc...; in contrast to media that is more specialized." Also defined by Webster's New World Dictionary: Media and Communications is mass communications, which means, "the delivery of information to large audiences via electrical and print media." Both of these terms are critically important in order to learn about mass media.

It is believed that Johannes Gutenberg began the revolution of mass media. It is widely held that he was the first to use movable type to print a large book. With the printing of *The Gutenberg Bible*, in 1455, the process of getting a message to reach a vast audience in a shorter period of time and at a reduced cost began. Soon after Gutenberg's press gained fame, governments used the power of the press and also attempted to restrict its use. The attempt to control the printing press was unsuccessful and it continued to reach even greater audiences. Tocqueville considered it incredible, when he commented in the early 1800's that newspapers had the extraordinary power to plant the same idea in 10,000 minds a day.

Today mass media has grown to unimaginable levels and can connect with anyone anywhere in the world. By using it's many forms of communication: television, radio, newspapers, magazines, advertisements, books, music, and movies; mass media has been able to create a smaller world, a global village. As newspapers communicate ideas with hundreds of thousands of people in a day, television is communicating an idea in hundreds of millions of minds instantaneously. All of this mass communication creates a world in which everybody is able to know about anybody and anything at anytime. Like a village: where you know what all of your neighbors are doing; mass media has created a global village uniting the world into a small intricate community.

There are many positive uses of the forms of mass media. One of the most common uses is to present an objective look at our world. The news we read, see, or hear about helps us to educate ourselves and others about what to do or not to do. When a person is informed by media about the trends of the New York Stock Market, they are able to make rational choices about their investment's. When advertisements state their sales and discounts, they are allowing the consumer to shop effectively. Many forms of media help us to relieve stress, such as television sitcoms, movies, or music. Mass media informs us of changes that will affect our lives. The purpose of the positive uses of mass media is not to hurt or slander others, but to present either objective and truthful information, or an insightful perspective of society.

There are also many negative uses for mass media. Yellow journalism, propaganda, and manipulative information are some of these negative uses. Yellow journalism is sensational reporting that attempts to attract people to it's product. Tabloids, both in newspapers and on television, are the most common form of yellow journalism. By twisting and stretching a fact, this form of media presents erroneous, extremely emotional, and often illogical stories to the public. Propaganda is another form of a negative use of mass media. Negative propaganda: misinformation used to influence the emotions of people: can be harmful when used. In World War II the Nazis used negative propaganda to manipulate their followers into believing the Nazi doctrine. This led to the deaths of over six million jews and countless others. News that is biased or 'slanted' preys on people's emotions and ignorance. Slanted news does not present an objective view or commentary, but a deceiving and untruthful one. News, slanted or not, can become harmful to society if it desensitizes culture. When problems or troubles are accepted by people, and not challenged because of their negative affects, they have become desensitized towards the issue and have accepted the occurrence as normal. Many people are worried about

BEST COPY AVAILABLE

violence on television. Studies have shown that as children see violence on television, they accept this violence as common and normal. Many people are worried that as children accept violence, they will act with violence. This negative information, and desensitization is harmful to the progression and development of our race. The last example of negative uses of mass media is that of advertising. Advertisements can be positive or negative uses of mass media. Advertising can play to a person's emotions; it can be truthful, it can be comparative, it can be hurtful, or it can be untruthful and twisted advertising. The last two are examples of negative uses by advertisers of mass media. When a political candidate urges people to vote for them because he hurts the other candidate with untruthful or deceitful statements, the candidate is using negative advertising. When a weight-loss food chain twists or distorts the facts about their foods or success rate in order for the consumer to buy their goods, they are using negative advertising. These are many forms of the negative uses of mass media. These are not used to present accurate or credible information to the public, but to deceive and lie to society about facts.

There are many influences on mass media. Often it is influenced by those who provide the sources for mass media. It is the editors and journalists who influence the content and coverage of news stories. It is the authors who influence the content of their books. The consumer often influences advertisements by their purchasing trends. The influences can often be exclusive based on wealth, power, or sex. The news industry has been very male-dominated. Thus the coverage of news stories has been very influenced towards 'male issues'. Often in the history of mass media, the issues that have been discussed or given recognition have been influenced by the rich and/or powerful. When a movie, television show, or book was disliked by an influential person, the person often had the opportunity to prevent and change the outcome of the project. There are many influences which affect mass media both for the better and the worse.

I use mass media for many of my daily activities. When I read, I am using a book, newspaper, or information pamphlet, which are all forms of mass media. When I watch television news, sitcoms, or dramas, I am using a form of mass media. When I relax to music I am using a form of mass media. Many of the things that I use and need, including information, and textbooks, are provided by mass media.

Today's society is greatly influenced by the forms of mass media. The pictures offered by mass media of the soldiers in Somalia affected the opinions of the American people. Our emotions and decisions are constantly being affected by both the negative and positive forms of mass media. As we integrate technology into our lives and create a global village, we require mass media to help us understand and communicate with the neighbors around us.

APPENDIX B

R.O.P.E. First Semester 1991-92

R.O.P.E. First Semester 1991-92

Wednesday, September 4	Continue writing the autobiography in class
Friday, September 6	Put the final polish on the autobiography
Monday, September 9	Introduce Letters of Recommendation and Resume and cont. auto
Wednesday, September 11	Introduce Science section, write coursework in class and cont. auto
Wednesday, September 13	**Autobiography due** – write technology essay and computer testimonial
Monday, September 16	Physical and Life Science in class work
Wednesday, September 18	**Science 1 page explanation due** (Phy. & Life) – Environmental explanation
Friday, September 20	**Letters of Recommendation due** – write science environmental/comp/tech
Monday, September 23	**Technology and Computer Use essays due** – begin employment resp.
Wednesday, September 25	**Environmental essay due** – Introduce current events & write employment responsibility
Friday, September 27	**Employment responsibility due** – begin bibliography and reading reports
Monday, September 30	Begin Physical data page and write book reports
Wednesday, October 2	**Reading list due** – begin Fine Arts
Friday, October 4	**Book Reports due** – introduce Mass Media
Monday, October 7	**Physical Data page due** – Ethics intro – continue writing
Wednesday, October 9	**Mass Media due** – continue writing other sections
Friday, October 11	**Fine Arts due** – continue ethics and other sections
Monday, October 14	**Current Events due** – Human Relations intro & continue writing
Wednesday, October 16	**Ethics due** – continue writing other sections
Friday, October 18	**Human Relations due** – begin Multi-Cultural Awareness
Monday, October 21	**Science Section due** (all parts) – continue with other sections
Wednesday, October 23	**Multi-Cultural Awareness due** – talk over organization, contents, etc.
Monday, October 28	**PORTFOLIO IS DUE – ABSOLUTELY NO EXCEPTIONS**
Wednesday, October 30	Give out evaluations
Monday, November 8	Introduction to Second Quarter
Wednesday, November 10	Picking a Topic (Lecture and Lehman handout)
Friday, November 12	Outlining and Research Notetaking – topics declared
Monday, November 15	Bibliography, Footnotes & Readers Guide Presentation (Sue K.K.)
Wednesday, November 17	Our Library with Sue to begin basic references
Friday, November 19	Parkside Library – bring ID & $5.00 for library card 8-10:30
Monday, November 22	Research question and answers (expl. opening thesis, etc.)
Wednesday, November 24	Eat Turkey and help Judy!
Friday, November 26	Vacation day – should be spent at the library doing research
Monday, November 29	Check on progress **GRADE** notes, introductory paragraphs & bibliography
Wednesday, December 1	RPL walk over 9:15 and 12:15
Friday, December 3	Introduce demonstrations – show vcr – sign up to present demos
Monday, December 6	MATH – 3 demonstrations (Madden Peak expl.)
Wednesday, December 8	SCIENCE – 3 demonstrations
Friday, December 10	GOVERNMENT – 3 demonstrations & handout
Monday, December 13	MULTICULTURAL – & **GRADE** – ½ completion grade
Wednesday, December 15	GEOGRAPHY – 3 demonstrations and handouts
Friday, December 17	FINE ARTS – 3 demonstrations and questions
Monday, December 20	US HISTORY – 3 demonstrations and handout (top 100)
Wednesday, December 22	HUMAN RELATIONS – 3 DEMONSTRATIONS & **GRADE** – ¾ completion grade
Monday, January 3	PERSONAL GROWTH & PROFICIENCY AREAS – 3 demonstrations
Wednesday, January 5	MASS MEDIA – 3 demonstrations
Friday, January 7	ENGLISH – demonstration, book reports and overall summary
Monday, January 10	Current Events – 3 demonstrations
Wednesday, January 12	Final Question day
Friday, January 14	**ALL RESEARCH PAPERS DUE IN CLASS – NO EXCEPTIONS**
Monday, January 17	Evaluations
Wednesday, January 19	Last Day

The College Board's Pacesetter Mathematics Program: Sommerville High School
May 10, 1995

SOMMERVILLE

THE COLLEGE BOARD'S PACESETTER MATHEMATICS PROGRAM: SOMMERVILLE HIGH SCHOOL

Introduction

In 1993-94, the College Board piloted the first of its *Pacesetter* programs in volunteer schools across the country. The College Board has developed these high-school level "integrated programs of standards, teaching, and assessment" in mathematics, English, and Spanish, and is in the process of developing programs in science and world history. The College Board views these Pacesetter programs as one element in its "push-pull" strategy of education reform, designed to encourage all students, particularly minority and disadvantaged students, to take rigorous academic courses during secondary school.

This case study describes how the Pacesetter mathematics program is being implemented in one Maryland high school. Sommerville High School is a large public high school located in the Maryland suburbs of Washington, D.C. In 1994-95 the school served 2641 students in the ninth through twelfth grades. These students were largely African-American (89 percent), with other students being white (8 percent), Asian-American (2 percent), and Hispanic (1 percent). The community the high school serves is primarily middle class, and about 14 percent of students qualify for the Department of Agriculture's free or reduced price lunch program. The school's daily attendance rate averages about 90 percent, and the annual dropout rate is low, at about 2 percent. Sommerville High School's staff includes 172 teachers, 8 administrators, 8 guidance counselors, 1 librarian, and 14 support staff members. Of the 172 teachers, 21 are mathematics teachers.

Sommerville High School participated in the Pacesetter Math pilot program during the 1993-94 school year. In 1994-95 Sommerville continued to participate in Pacesetter, offering Pacesetter mathematics to about 180 students that year. The Pacesetter course is taught by only three of the school's 21 math teachers and reaches only a small fraction of the school's students. Though the program seems to have become quite popular among participating students and their teachers, any impact the school's participation in Pacesetter may exert on mathematics teaching and learning outside Pacesetter classrooms remains, at this point, quite limited in scope.

Pacesetter assessment is, as the College Board's description of its programs suggests, integrated with the program's curriculum and instructional approach, making it difficult to isolate the program's assessment component for consideration. Therefore, this case study explores the implementation and effects of the Pacesetter program, including the project work, task sets, and other performance-based assessments, as a whole at Sommerville High School.

Participants

Members of the Sommerville High School community who participated in this study are identified in Exhibit 1. In addition, Lola Greene, the College Board's Pacesetter Project Manager, was interviewed by telephone.

EXHIBIT I

Study Participants

- Supervisor of Math (district office)
- Principal, Sommerville High School
- 4 mathematics teachers — all 3 teachers who teach the Pacesetter course were interviewed, as was the chairman of the mathematics department
- 12 students enrolled in Pacesetter
- 2 parents of students enrolled in Pacesetter

Observations

In addition, two Pacesetter classrooms were observed during 45-minute academic periods.

Pacesetter

The Pacesetter programs were developed to complement the College Board's already existing *Equity 2000* program. Together, the two programs form the core of what the College Board calls its "push-pull" reform strategy: Equity 2000 is designed "to 'push' students, particularly minority and disadvantaged students, into more demanding academic preparation by requiring them to take pre-algebra in the eighth grade, followed by algebra and geometry."[1] Pacesetter, in turn, is designed to "pull" students "toward a goal of high standards of achievement for all students before graduating from high school."[2]

The Pacesetter programs in mathematics, English, and, eventually, science are designed to be fourth-year "capstone" courses that pull together for students the subject matter and skills they have learned during their elementary and secondary school careers and help students realize the applications of these skills in "the real world." In contrast, Pacesetter Spanish is a third-year Spanish course, and the Pacesetter world history course is intended primarily for tenth-grade students.

All Pacesetter programs incorporate three central components:

- *Statement of Standards* — Course-content outlines include statements of the specific knowledge and skills students are expected to have mastered upon completion of the course. Teachers and students share the same understanding of what must be taught and what must be learned.

---

[1] *Facts about Pacesetter*. The College Board, 1994.

[2] *Facts about Pacesetter*. The College Board, 1994.

- *Teacher Preparation* — Pacesetter teachers participate in a variety of ongoing activities designed to support their teaching, including summer and midyear institutes, scoring sessions, electronic communication links, 800 hotline assistance, and access to publications about successful teaching activities.

- *Assessments* — Pacesetter teachers and students make use of a range of assessments, including short-answer questions requiring students to construct responses, performance-based tasks, essays, projects, and portfolios. Students are expected both to maintain a journal and regularly contribute to a portfolio illustrating their work in the course. This range of tasks allows students to demonstrate what they have learned in a variety of ways. Assessments embedded in the course are designed to give students feedback on how they are doing and teachers information to help them monitor and modify instruction. Culminating assessments are used to give information on achievement to students and their teachers; to evaluate the overall performance of students, classes, and schools in achieving standards; and to provide employers and colleges with a range of information about students.[3]

Purposes and Objectives of Pacesetter Mathematics

As has been described above, the Pacesetter mathematics program is designed to be a "capstone" course in math, taken primarily by students in their senior year of high school. The College Board says the curriculum is between an algebra II course and a pre-calculus course, and the course prepares students either for further academic study (typically a first-year calculus course) or for entry into the workforce. The course is intended to support two of the national education goals: "demonstrated competency in challenging subject matter" and being "first in the world in math and science;" it is also designed to reflect the standards put forth by the National Council of Teachers of Mathematics.

Specific academic objectives of the Pacesetter mathematics program include teaching students how to:

- develop models from data for situations based in linear, polynomial, exponential, logarithmic, trigonometric, and recursive functions;
- delineate the numerical and graphical differences in the nature of linear, polynomial, exponential, logarithmic, and periodic functions, noting the basic properties of each form;
- use a function-based model to predict changes in one of the variables involved in a given mathematical model knowing the nature of changes in the other variables;
- describe the nature of a mathematical model by using an example, illustrating how a model helps interpret and resolve a problem situation;

---

[3] The College Board, *Facts about Pacesetter*, 1994.

- investigate a mathematical model for a given context to determine how the values or processes involved in the situation can be optimized;
- understand how mathematics can be a tool for understanding, appreciating, and applying knowledge in other disciplines, such as science, commerce, social science, and the humanities;
- communicate mathematical concepts and relationships via reading, writing, speaking, listening, and drawing; and
- reason clearly in situations involving numerical, spatial, graphical, or analytical information.[4]

The Development of Pacesetter

The College Board initiated development of its Pacesetter curricula as part of its ongoing contribution to education reform. Pacesetter mathematics was developed by a 12-member task force: the task force was chaired by John Dossey of ETS and included both high school and collegiate mathematics instructors. In addition, teacher representatives of the Mathematical Association of America and the National Council of Teachers of Mathematics participated throughout the development process. The task force's development work took about two years to complete. However, Ms. Greene emphasizes that Pacesetter Math remains a work in progress that the College Board will continue to modify based upon feedback received from the field.

Characteristics of Pacesetter Math

The College Board summarizes the approach of Pacesetter mathematics as follows:

> Pacesetter mathematics places emphasis on mathematical modeling, data analysis, prediction, and optimization. The course will employ students' previous knowledge of algebra, geometry and elementary probability and statistics. Technology is employed as a tool for bringing about a deeper understanding of mathematical structure and relationships. The unifying focus of the course is the concept of function, and the course itself is organized into six major units of study that develop understanding of the nature of functions, their properties, and their applications. Further, these six units are designed in such a way as to fit with current textbook materials and to articulate with post secondary mathematics programs, be they university or vocational in nature.[5]

The six units into which the Pacesetter mathematics curriculum is organized (in the 1993-94 pilot there were five units; one additional unit was added in 1994-95) include: (1) modeling with linear functions; (2) modeling with exponential and logarithmic functions; (3) modeling with power, polynomial, and rational functions; (4) modeling with trigonometric functions; (5) modeling with matrices; and (6) modeling with more general functions.

---

[4] The College Board, *Pacesetter Mathematics*, May 4, 1994, p. 3.

[5] *Pacesetter Mathematics*, The College Entrance Examination Board and Educational Testing Service, 1994, pp. 3-4.

Each unit comprises a series of instructional activities, including projects, essays, performance-based tasks, reflective journals, and short-answer quizzes. "Task sets" comprise the heart of the units, and each unit may have several "task sets." These task sets, which frequently include problems students work on in small groups, generally take students several weeks to complete. Teachers determine how they use and score these assorted task sets and other activities.

A central component of the Pacesetter mathematics program is the use of graphing calculators. All schools participating in the program are expected to make this technology available to students.

Finally, an end-of-the-year assessment is given to all students enrolled in Pacesetter. This assessment, which includes two 90-minute segments and must be administered by teachers during a specified week in April, is conducted by the College Board. Three dimensions are included in the assessment: mathematical knowledge; applied problem solving; and communication in the language of mathematics. Pacesetter teachers come together in early May to score the assessments, using rubrics also developed by the College Board. The College Board then provides teachers with the scores of their students prior to the end of the school year.

The College Board intends to use the results of the assessment to track trends in student performance over time; in the long run, the Board hopes to do some comparative studies as well (for example, comparisons of performance in AP Calculus courses between students who have completed the Pacesetter program and students who have not). However, the culminating assessment itself is still evolving rapidly, and Ms. Greene emphasizes that the College Board is using results from the 1993-94 and 1994-95 assessments cautiously and that the assessment is, and will remain for the foreseeable future, low stakes for both students and teachers.

Pilot of Pacesetter Math

Pacesetter math was the first Pacesetter program to be piloted. The pilot year was the 1993-94 school year, and 15 schools in 10 school districts participated in the pilot. The College Board funded the pilot, and so the only cost to schools and districts was release time provided to participating staff to attend various professional development activities.

The pilot provided the College Board with information about the content and implementation of Pacesetter Math, including problems that Pacesetter participants face in using the program. Information collected during the pilot is being used to modify the program. Specific lessons learned from the pilot include:

- The Pacesetter Math program provided a rigorous course of instruction for students.
- On the pilot culminating assessment, students experienced the most difficulty with tasks that required them to "communicate in the language of mathematics."

- The pilot included a "broad range" of students, and only a few of them were found to be "Pacesetter proficient" on the pilot culminating assessment.

- The College Board would need to modify Pacesetter materials for both teachers and students. A "common task" to be included in the student's portfolio was added to each unit.

- Although Pacesetter mathematics was designed to be taught in conjunction with a textbook, teachers were not using textbooks because no compatible textbooks were available to them.

- Pacesetter materials needed to incorporate more "student-friendly" instructions.

- Cooperative learning is a valuable and challenging instructional technique, for both students and teachers; highest-performing students have the hardest time with cooperative learning activities.

- The pilot culminating assessment was too long and needed to be more closely aligned with Pacesetter curriculum tasks. Both of these problems were addressed in the 1994-95 assessment, which, according to Ms. Greene, most Pacesetter teachers agreed was a much fairer assessment than the previous year's.

Resource and Professional Development Support

The development of Pacesetter was funded by the College Board and the Educational Testing Service (ETS); the pilot implementation of the math program in 1993-94 was also financed by these two organizations.

However, beginning in 1994-95 (for the mathematics program) schools using the Pacesetter program were required to pay for the curriculum and the support provided by the College Board and ETS: in 1994-95, schools paid the College Board $30 per student enrolled in the Pacesetter math program.[6]

As "teaching" is one of the three elements of the integrated programs, the College Board and ETS have provided substantial support to schools and teachers involved in the program. This support has taken on five forms:

- ***Summer institutes and mid-year "refreshers"***. Teachers are introduced to Pacesetter during the week-long Summer Institute. During the 3-day mid-year refresher, teachers are able to present the problems they have encountered and discuss solutions (how to develop appropriate assessments is a frequently discussed issue and one that poses problems for many teachers). Participating districts help defray the costs of the institutes. Most Pacesetter teachers attend both the summer institute and the mid-year refresher.

---

[6] However, this amount does not represent the full cost to the College Board of running Pacesetter; scoring the culminating assessment alone cost the College Board and ETS $67 per student in 1994-95. Reconciling the costs of the program with the amount school districts are able to pay to use the program represents an important challenge to the College Board as it continues to develop and market Pacesetter. (In the long run, the College Board aims to recoup the development and operating costs associated with Pacesetter, not to make a profit.)

- ***Assessment "reading" sessions***. This "major component of professional development" focuses on scoring the culminating assessment. Teachers are trained using the model the College Board employs with scorers of AP exams.

- ***On-site technical assistance***. During the pilot year, College Board representatives visited every Pacesetter classroom, both to collect information about how teachers and students were reacting to Pacesetter and to provide professional support to teachers. During the 1994-95 school year, schools could request this sort of assistance. Visits from "mentor-observers" typically lasted two days, and approximately 20 percent of schools participating in Pacesetter Math in 1994-95 requested and received this type of assistance.

- ***Electronic communications links***. The College Board maintains an electronic bulletin board called the "College Board Connection;" 800 number hotlines are also available by subject area.

- ***Publications***. The Pacesetter newsletter keeps teachers apprised of developments and describes successful teaching strategies.

Interaction with Other Reforms

Pacesetter is one of several programs developed and administered by the College Board to support education reform and to promote high quality academic programs in schools. The two important College Board programs Pacesetter joins are the Advanced Placement (AP) program and the Equity 2000 program.

The College Board's AP program has been in place for many years. It promotes the teaching of college-level courses to high school students; most U.S. colleges and universities accept passing scores on the AP examinations for credit toward graduation and placement into higher-level courses. Pacesetter programs, on the other hand, teach advanced high school-level, not college-level, courses to students. However, Pacesetter, the College Board suggests, was in many ways modeled on its successful AP programs: both programs (1) are designed to raise standards and achievement levels, (2) provide a syllabus, course outline, professional development, networking opportunities, and examinations developed by and for teachers, and (3) have culminating examinations that contain performance-based elements.[7]

Pacesetter Mathematics is also related to the College Board's Equity 2000 program, providing the "pull" to its "push," as described above. Equity 2000 promotes algebra and geometry coursetaking among minority and disadvantaged students, and Pacesetter takes students further in their mathematics studies.

---

[7] *Facts about Pacesetter Mathematics*, College Board, 1994.

District Context

The district in which Sommerville High School is located has long been a participant in the College Board's Equity 2000 program. Thus, when the district's mathematics area supervisor learned of the new Pacesetter program, she pursued the idea of joining the pilot. She says that the type of integrated approach to instruction and assessment is the direction she wants to move mathematics teaching and learning in the district.

Sommerville High School was an obvious choice for the pilot school, says the Math Supervisor, because the school's mathematics department chair was also very enthusiastic about Pacesetter. The principal was also impressed by the program and promised his support as well.

The district agreed to provide the necessary financial support ($30 per student enrolled in the program) to continue participation in Pacesetter in 1994-95 (participation in the pilot was funded by the College Board and ETS), and the district has committed itself to fund Pacesetter at Sommerville through the 1995-96 school year. However, the Math Supervisor says that the program is too expensive for the district to be able to support its introduction into other high schools. "It is time to start talking with the College Board about the next steps with Pacesetter because it is too expensive the way it is now," commented the Math Supervisor.

School Context

Sommerville High School provides a wide range of academic opportunities to students. It is a magnet school for the district, serving students gifted in the visual and performing arts. In addition, the school operates a "university high school," a program for gifted students who are not enrolled in one of the district's two "flagship" schools for the gifted. Finally, the school also offers courses through the International Baccalaureate program, and it has a strong vocational education program as well. Within this wide range of offerings, the school's academic program is primarily college preparatory, and the vast majority of the school's graduates go on to pursue college or other postsecondary education.

As was mentioned above, Sommerville High School was identified by its district for participation in the College Board Pacesetter program both because of its ongoing participation in the College Board's Equity 2000 program and because the school's mathematics department chair was enthusiastic about experimenting with the program. Sommerville was the only high school in the district to participate in Pacesetter from its outset, though in the 1994-95 one other district high school began to participate as well. It was one of 15 schools across the country to participate in the 1993-94 pilot of Pacesetter mathematics, and one of approximately 50 schools to be involved with the program in 1994-95.

Unlike most other schools participating in Pacesetter math, Sommerville High School has chosen to target the program at students who have struggled in their previous math courses.

Though these students have completed a course in algebra II, most of them earned grades of "C" or "D," and teachers considered them to be at risk of not taking additional math courses during their high school years. Teachers say that the course is well-suited to these students, even though it was not specifically designed for them.[8]

Impact of Pacesetter Mathematics

All things told, Sommerville teachers, students, and parents interviewed for this study have favorable opinions of Pacesetter Math. Interviewed students are universal in their praise of the course, and teachers, though they have some concerns about the program, believe it is truly an effective approach to teaching students mathematics. Sommerville's principal is very supportive of the school's three Pacesetter teachers; in particular, he agreed to their request that they be provided with an additional planning period each day, scheduled so that the three teachers could meet together to discuss instruction and assessment techniques. Finally, parents — both the two participating in this study and, reportedly, others as well — have responded positively to Pacesetter.

Impact on Teachers

Three math teachers at Sommerville High School teach the Pacesetter program. Two of the three teachers were involved in Pacesetter during its 1993-94 pilot; the third teacher began teaching Pacesetter during the 1994-95 school year. The two involved from the outset were tapped by the district's mathematics area supervisor and their department chair to participate because of their experience with the Equity 2000 program.

The three teachers express primarily positive comments about Pacesetter. They say that the approach to teaching mathematics was already compatible with their teaching styles and philosophies. However, they also acknowledge a real shift in their teaching resulting from participating in Pacesetter.

These three teachers say that the program provides, in the words of one teacher, "a good class to help students go out in the world. It is a good place for students to straighten out misconceptions about math, and they develop good social skills through the group work, too." The teachers also said they like the fact that the program draws on students' skills in the areas of reading, reasoning, writing, and communication, as well as math skills.

The teachers did, however, express some reservations about the Pacesetter program. These included:

---

[8] Teachers encourage individual students who they think would benefit from the course to enroll in it. For the 1994-95 school year, however, the majority of the students enrolled in the program were enrolled in the course by choice. The class called "trig analysis" was canceled, and all students who had signed up to take that course were enrolled in Pacesetter instead. As will be explained below, students interviewed for this study claim that they, and most of their Pacesetter classmates, were unhappy at the beginning of the school year about the placement but have since become converts.

- Although the College Board presents the program as a math course bridging algebra II and pre-calculus, Sommerville High School teachers found the course to lean more toward the pre-calculus. However, they suggest that, for most students, the course is not adequate preparation for a calculus course.

- The curriculum, as presented by the developers, does not adequately incorporate the teaching of "math skills" into the task sets. Teachers said that, while last year they concentrated on having students complete the task sets, this year they had opted to cover math skills in more traditional ways before introducing the task sets. They said that this approach — combining the applied work conducted on task sets with the teaching of skills — has been more successful with students, who now understand and retain the material better.[9]

- The curriculum is too optimistic for a year-long class. The two teachers who taught Pacesetter during its pilot said that they did not get through all five units of the course, and they are certain that they will get no further this year, though the College Board has added a sixth unit to the program. One teacher said, "I don't even know what the last unit is. Last year there were five units, and this year they added one, so now there are six. But last year most people didn't get through five. Go figure."

- The College Board has not yet identified which types of students may benefit most from Pacesetter mathematics. Teachers suggest that "the College Board is trying to make something for everybody."

However, Sommerville's teachers emphasize that these complaints about Pacesetter Math are minor, especially in light of the benefits they believe the program offers their students.

Finally, one math teacher who does not teach the Pacesetter program commented on how Pacesetter is affecting other teachers and classrooms at Sommerville High School. He suggests that the instructional approach is very promising, and he noted that he had adopted a few ideas from the Pacesetter teachers in his classrooms. For instance, he says he has used some Pacesetter ideas about calculator applications in his math classes. However, he also expressed doubts about coverage of topic areas using the Pacesetter teaching approach. For example, he said, "I'm sure the Pacesetter approach would be an effective approach to teaching calculus, but I doubt it's a time-effective approach." He went on to suggest that, "Pacesetter is great for the kid who is willing to meet you half way. The kid who's not motivated at all by a problem probably won't be motivated by Pacesetter either."

Teachers' reactions to the summer institutes and mid-year refreshers. Two of Sommerville's three Pacesetter teachers have participated in two Pacesetter summer institutes put on by the College Board. The first of these, which preceded the pilot implementation of Pacesetter, was held in San Diego and lasted about 10 days. This program was designed to help teachers learn how to teach the Pacesetter math course. The two teachers who attended said that the institute was very valuable to them as teachers, but that they were surprised by the advanced level of the curriculum; they had believed that the Pacesetter program would be below the level of a pre-calculus course, but that in reality it was essentially a pre-calculus course. The teachers also acknowledged that they felt some resentment that the institute was led by instructors from "elitist high schools;" teachers from such schools, Sommerville teachers said, were not in touch with the types of issues and challenges confronting teachers and students in less privileged schools (including students who are not prepared to do the type of work required by Pacesetter).

---

[9] Note, however, that the College Board intended the Pacesetter curriculum to be supplemented by a standard textbook. The unavailability of such a textbook poses problems for these teachers, though they have taken their own steps to remedy the problem.

These two teachers also participated in the first "mid-year refresher," held in Dallas in January 1994. This support from the College Board, the teachers said, came too late. The College Board apparently agreed with this feedback from participating teachers, for in the 1994-95 school year, the mid-year refresher was pushed up to December 1994. A summer 1994 refresher was also conducted, this time in Charlotte, North Carolina. (Teachers did not provide any specific comments about their participation in these events, though the two Sommerville teachers participating in Pacesetter from the beginning attended each of the activities.)

Teachers' reactions to the end-of-the-year assessment and to participating in scoring sessions. All students enrolled in Pacesetter take an end-of-the-year assessment developed by the College Board. Sommerville Pacesetter teachers said that the assessment administered in 1994 was "a rude awakening for the College Board." Students, they said, could not answer the problems included in the assessment. In these teachers' opinions, the assessment tasks poorly matched the Pacesetter curriculum and, thus, the assessment was not an accurate measure of the learning taking place in Pacesetter classrooms. They said that the revised assessment administered in spring 1995 had been "totally rewritten" from the previous year and that, though challenging to most students, it was better aligned to the Pacesetter curriculum and more realistic in what it asked of students.

Two Sommerville Pacesetter teachers have participated in the scoring sessions (or, as the College Board calls them, "reading sessions") convened by the College Board to score the end-of-the-year assessments. Teachers said that during these sessions the scoring rubrics (developed by the College Board and ETS) are explained to them, and they then score assessments using them. Overall, the two teachers found their two reading experiences to be valuable. Though they had reservations about some of the rubrics (they felt some of the rubrics should have had more points to differentiate among levels of performance), teachers said they learned from these sessions, both about Pacesetter and about the application of rubrics (which teachers develop, though informally, for use in their Pacesetter classes). In the words of one teacher, "Our rubrics are like theirs, but they're better at it than we are. The training session on the application of rubrics helped teachers realize how crucial consistency is in scoring."

Finally, teachers said that some Pacesetter teachers who participated in the reading sessions complained about the one-week window in April during which the College Board required them to administer the assessment. For some schools, this week directly followed the spring break, and teachers felt it was unfair to require them to give their students an assessment right after a vacation.

Impact on Instruction

The three Pacesetter teachers at Sommerville High School say that their instructional approach has evolved since their involvement in Pacesetter and that this evolution has carried over into their other math classes as well. One important difference between Pacesetter math classes and other math classes is the heavy use of technology (in the form of graphing calculators) in the Pacesetter program. Teachers say that the calculator is a very effective teaching tool in their classrooms. "Students no longer ask 'what if?' so much. They can find out 'what if' on their own using the calculator. We use the 'discovery method' much more now." One teacher said, "The graphing calculator is going to be the great equalizer in math for kids. These kids normally wouldn't do this kind of work." A description of a Pacesetter classroom in action appears in Exhibit 2.

The Pacesetter program has further driven instructional change by encouraging more group work. Teachers say they have observed in their students improved social skills, and they attribute this growth, to a great extent, to the frequent group Pacesetter assignments. On teacher noted that he had become more adept at assigning group work over the course of the school year, especially with respect to achieving satisfactory balances in group dynamics.

The two Pacesetter teachers who participated in Pacesetter's pilot noted that they had to make dramatic adjustments to the ways in which they assess students' achievement. When they began to teach Pacesetter, they used tests very much like the ones they used in their other classes. They quickly discovered, however, that these tests were not compatible with the Pacesetter approach to teaching, and they began to construct more performance-based assessment tasks. (Note that these teachers' experience resonates with the College Board's finding that teachers returning to mid-year refresher courses frequently needed help in designing appropriate assessments.)

One Pacesetter teacher also noted that he had taken the instructional techniques he had developed for teaching Pacesetter and adopted them in his other math classes as well. For instance, he now emphasizes applications more in calculus courses, and he asks his algebra students to keep math journals.

Impact on Students

All twelve students interviewed for this study were enthusiastic about Pacesetter. More than any other comment, students suggested that the applications of math they understand through their work on task sets are valuable in "real life." One student said that math is more enjoyable this way: "When you get algebra, stuff with variables, you don't think you'll use it. But in Pacesetter we used it with real life situations." Another student said, "You would never think it would take a function to get an interest rate. I want to go into business, and that is relevant to me." One student's mother said that her daughter, after working on interest rate functions in her Pacesetter class, had sat down and developed a function to calculate how much money she would need from different sources to fund her college education. She is now bound for Boston University. (All students interviewed expected to attend college directly upon completion of high school. However, unlike Sommerville's Pacesetter students — about 80 percent of whom are seniors — most of these students were juniors.)

EXHIBIT 2

An Example of Pacesetter Instruction

During a 45-minute class period, Mr. Kearney (whose students all call him Sam), taught a lesson on hyperbolic functions of the form f(x) = 1/x. Putting the equation f(x) = 2/(x-1) on the board, he proceeded to pepper his students — 8 boys and 7 girls — with questions: Is this a decreasing function? What does the graph of the function look like? What is the name of line where the graph breaks? At what value of x does the asymptote appear? When students volunteered answers, Sam followed up, asking why they knew that that was the answer or how the current function was different from other particular functions. When the answers were slow in coming, he pressed students, reminding them of examples covered earlier in the week.

After discussing the properties of the function under study, Sam asked the students to turn on their graphing calculators. Using a liquid crystal display to project the output of his calculator, Sam illustrated what the function looked like. He coached students as they correctly programmed their own calculators to represent the function. Then, using the calculators, the class went on to make various adjustments to the equation, first to see what would happen to the graph and values of the function if certain changes were made, and then, based on the information collected from the first graphs, to develop and test hypotheses about what would happen to the graph and values of the function if certain changes were made to the function.

Most students said that they like the group work Pacesetter often involves (teachers noted, however, that the very best students in their classes liked group work less; teachers attribute this response to the students' sense that they end up doing a disproportionate share of the work). Several students also said that they like that there is no "one right way" to solve problems in Pacesetter. "Everybody uses different ways of finding out answers. It makes it easier to understand," said one girl. Another student said that the course has been a real "confidence booster" for her; she felt that she liked and understood math for the first time.

Students' one complaint about Pacesetter was that they did not like its name. "It sounds like it's remedial or something. Why'd they give it such a stupid name?" several students asked.

Impact on Parents

The two parents interviewed for this study were both well informed about Pacesetter and were extremely pleased with the program's impact on their daughters. Both parents said that, though their daughters had always been good students, they had struggled with math; one girl was an "A" student in her other classes but a "C" student in math, while, according to her mother, the other girl earned A's in math but had limited comprehension of what she was being taught.

Both parents said Pacesetter had turned around their children's understanding of mathematical concepts, their attitudes toward math, and their self-esteem with respect to learning math.

One of the parents, who is a biology teacher at Sommerville High School in its International Baccalaureate program, also said that she had, over the past two years, noticed a difference between her students who were enrolled in Pacesetter and those who were not. She found that her Pacesetter students already knew the mathematical applications that she had always had to teach her biology students. In this woman's words, "As a parent, I give Pacesetter a 10 out of 10. As a teacher, I give it a 12 out of 10."

Future Plans

Sommerville High School anticipates continuing its participation in the Pacesetter mathematics program. The school and district would like to experiment with using individual Pacesetter units with students enrolled in non-Pacesetter classes. However, as of right now they do not see how they can do so, given the fact that the College Board charges the district per student for the use of Pacesetter curriculum. (Hence the math supervisor's comment that they will need to discuss "next steps" with the College Board.) For the time being, however, Pacesetter will remain a discrete math class at Sommerville High School.

Conclusions

Teachers, students, and parents alike express their confidence that mathematics learning has improved at Sommerville High School in Pacesetter classrooms. Though its impact rests almost exclusively in the classrooms of the three Pacesetter teachers, these teachers are sharing their experiences with their colleagues, some of whom are reportedly beginning to adopt some of the instructional and assessment techniques fostered by Pacesetter. Over time, it is possible that Pacesetter will become an important vehicle of reform in the ways in which teaching and learning takes place in Sommerville's mathematics classrooms.

The realization of this possibility, however, is dependent upon the district's continued funding of the program. Though the district is enthusiastic about Pacesetter and intends to maintain this level of funding, the Supervisor of Math voices concerns that participation according to the College Board's current requirements — a fixed cost per enrolled student — may become prohibitively expensive (and certainly precludes expansion of Pacesetter into other district schools). How the College Board addresses the issue of costs to districts participating in Pacesetter programs will surely have an important impact upon how far the effects of Pacesetter might reach.

The New Standards Project:
Ann Chester Elementary School
May 25-26, 1994

ANN CHESTER

THE NEW STANDARDS PROJECT: ANN CHESTER ELEMENTARY SCHOOL

Introduction

Two streams of ideas — from the New Standards Project (NSP) and the district's own Applied Learning Program — are transforming education at Ann Chester Applied Learning Center, a public elementary school in Fort Worth, Texas. Ann Chester Applied Learning Center's teachers and students are a self-selected group. To gain entrance into Chester's special community, teachers and students must apply to and be accepted by a school committee. (As of May 1994, about 250 children were on the school's waiting list.) The school's 18 teachers serve 380 students from K-5, who come from diverse backgrounds. About half of the students are white, 27 percent are African American, 20 percent are Hispanic, and 3 percent are Asian. Thirty-three percent receive free or reduced lunch.

Information sources for this case study include interviews with various key people and observations of assessment-related activities.

Participants

Interviewees are listed in Exhibit I.

Observations

Observations include an *Assessment Night*, a student-parent end-of-year conference focused on examining students' portfolios and in-class student participation on assessment tasks.

To gain a clear understanding of the school's involvement in the NSP and the district-based *Applied Learning Program* (ALP), it is necessary first to understand the structures of the NSP and the ALP. In this case study, we outline the NSP's efforts at developing a national assessment system and its system of recruiting participants, summarize the district's own reform program, and discuss Ann Chester's use of the NSP-driven assessment system as well as its own ventures into educational reform.

New Standards Project

Beginning with its inception in 1991, the aim of the *New Standards Project* has been to reinvigorate and revamp American education. In the NSP's own language, ". . . goal is to be a revitalized education system using assessment as a tool for transforming instruction and learning."[1] The program's concerns are the same as those being echoed in every hall of education — that American students are not thinking or writing as well as they should. The NSP program is being

---

[1] Simmons, W. & Resnick, L. (1993, February). Assessment as the Catalyst of School Reform. *Educational Leadership*, p. 12.

EXHIBIT I

Study Participants

- The district Director of Research and Evaluation
- The NSP site coordinator
- Chester's two teacher directors
- Four teachers
- A local NEA representative
- Three students from grades 3, 4, and 5
- Three parents
- One school board member

jointly directed by Marc Tucker, President of the National Center on Education, and Lauren Resnick of the Economy and the Learning Research and Development Center, University of Pittsburgh.

The crux of the NSP's work involves establishing content and performance standards and designing curricular, instructional, and assessment strategies. The NSP Board, which guides the formulation of content and performance standards and assessment strategies, is composed of representatives from the NSP's partner states and districts, and professional organizations such as the National Council for Teachers in Mathematics (NCTM), the American Association for the Advancement of Science (AAAS), and the National Council of Teachers of English (NCTE).

NSP Participants

In 1995, the NSP program listed 17 state and 6 urban district partners. Through its partner states and districts, NSP invited administrators and classroom teachers to help define content standards and develop the assessment system. Each partner's engagement with the project is generally coordinated by an on-site NSP coordinator, and participant teachers are responsible for developing assessment tasks and scoring rubrics and for conducting pilot tests in their classrooms. Teacher developed tasks are sent to an NSP committee (and teachers receive payment if the tasks are adopted for use by the NSP). Teachers also provide input into the content standards. Thus, grassroots teacher involvement at all stages — task development, administration, and scoring — is an integral component of the NSP effort.

The NSP Assessment System

The NSP assessment system is being formulated for grades 4, 8, and 10, and the fully articulated system will consist of student portfolios that will contain NSP recommended matrix-sampled tasks requiring extended responses, exhibitions, projects, and other student work. (The NSP tasks will be based on NSP content standards and evaluated using NSP performance standards.) The NSP portfolios will contain:[2] (a) work chosen by the district, school, teacher, and student; (b) prescribed projects and other extended learning activities; and (c) responses generated by NSP matrix examination tasks.

The NSP piloted a number of its assessment tasks in 1992, 1993, and 1994, in its partner states and districts. Classroom teachers and content area specialists scored these pilot tests using established scoring rubrics. The NSP anticipates the first valid, reliable, and fair exams will be available for use in mathematics and in English Language Arts by 1994-95; in Applied Learning by 1995-96; and in science by 1996-97. Portfolios will be in use by 1994-95.

The NSP Training System

To facilitate adoption of its assessment system, the NSP is helping to build professional capacity at the local level. Teams of 12 teachers from participating states and districts attend NSP assessment task and scoring development, curriculum development, and portfolio design sessions. After a number of training sessions, teachers are designated as *Senior Leaders*, and they, in turn, offer professional development in the same activities to other teachers in their districts and states.

NSP Funding[3]

In 1991, the program began with a $2.5 million donation fund from the Pew Charitable Trusts and the John D. and Catherine T. MacArthur Foundation. In June, 1992, the same two agencies donated a combined $8.5 million, and the Federal government funded specific projects. In subsequent years, partner states and districts contributed additional money.

District Context

One of the NSP district partners is Fort Worth Independent School District (FWISD). Located in north-eastern Texas, the district is fairly urban and operates 120 schools, all of which are governed by School Based Management (SBM) teams. Because of its involvement with the NSP and the ALP (described below), the district is regarded in Texas as a leader in education reform.

In addition to its own programs, the district must conform to certain state strictures. The district must adopt the *Texas Essential Elements*, which are content standards outlined by the state. In math, however, the district uses the NCTM standards, which the state has not utilized. In addition, the district administers the mandated *Texas Assessment of Academic Skills* (TAAS) to every child in grades 3 through 11.

The TAAS is "high-stakes," as high-performing schools win cash rewards and low-performing schools are assigned peer reviewers (if the state decides not to assume direct control of the school). The state legislature, though, is discussing the possibility of introducing performance assessments into the state. (The district discontinued its own writing assessments in 1993-94, when the TAAS administration was moved from fall to spring.)

---

[2] Simmons, W. & Resnick, L. (February, 1993). Assessment as the Catalyst of School Reform. Educational Leadership, p. 13.

[3] The *New Standards Project*, p. 12.

NSP Partnership

The FWISD joined the NSP in June of 1991, barely after the NSP's inception. Since then, district teachers have been involved in developing assessment tasks and in piloting and adopting NSP assessment guidelines. District teachers piloted the NSP assessment literacy tasks in grades 4 and 8, in the spring sessions of 1993 and 1994, which were sent back to the NSP for scoring. (As of May 1994, NSP had released some of these literacy tasks, but no anchor papers and no pilot results had been released.) Ninety teachers from 36 schools are involved with NSP, and FWISD remits an annual fee to the NSP (the amount of which varies from year to year).

Applied Learning Program

The district is immersed in its home-cultivated Applied Learning Program (ALP). The program makes explicit connections between the classroom and the world of work, and advocates the use of performance assessments to evaluate student progress and achievement. District-wide, 151 teachers from 52 district schools are participating in the program.

The program coordinator, called the Writing and Reasoning Coordinator, conceptualized and launched the program in 1991, which began with an analysis of 3,000 jobs in the area. A team identified the types of skills required to perform the different jobs included in the analysis. Based on this process, jobs were placed on a continuum of difficulty. The team then delineated a profile of *Competent Adult Performance* for each level of difficulty.

The Applied Learning teachers use these profiles of *Competent Adult Performance* to inform their curriculum and instruction. Curriculum and instruction is, thus, driven by skills requirements of the adult world of work.

Resource and Training Support

The district's NSP Coordinator is instrumental in (a) procuring funds for NSP-related training, (b) disseminating the NSP information, and (c) shouldering myriad responsibilities associated with coordinating the NSP work at the local level. The Coordinator also is the director of the district's ALP (spending 25 percent of her time with the NSP and the rest with the ALP).

Business Resources Support

In Fort Worth, business leaders and school administrators have become education reform "bedfellows." Area businesses support the reform agenda — in particular, the ALP — through the *Community, Corporations, and Classrooms* (C^3) Initiative (an initiative that won the 1994 National Community Award). As part of C^3, for example, area companies (such as Lockheed, IBM, Motorola, and Bell Helicopter) will accommodate 7,000 middle school students for two-week internships in the summer of 1994.

Financial Support

The Sid Richardson Foundation has donated $1 million for the district's ALP, NSP, and Performance Assessment Collaboratives for Education (PACE) programs.[4] According to one program coordinator, private individuals, too, have donated generous sums to the various programs.

Training

The ALP program coordinator provides intensive training in instructional and assessment strategies to district teachers. One hundred and fifty District teachers, including 18 from Chester, have completed or are currently undergoing training. Two years and three summers of training qualifies a teacher as an *Applied Learning Teacher*. In addition to the Applied Learning Training sessions, the district also held 12 performance assessments scoring sessions during 1993-1994.

Impact of Assessment Reform

According to district officials, the relatively slight teacher opposition to performance assessments attests to the "gentle nature" of reform processes in Fort Worth. Champions of performance assessments, such as the director of *Applied Learning*, use persuasion rather than coercion to nudge teachers toward assessment reform.

At present, teachers volunteer to attend professional training in applied learning. The ALP director sees training the few teachers who "absolutely do not want to use performance assessments" as the next hurdle to overcome.

According to one School Board member, the district spends a ". . . fair amount of money on public relations." The *Applied Learning* Schools, for instance, hold parent-staff meetings on a regular basis, and the ALP regularly briefs the business community and gives presentations on reform issues.

The community, thus far, has voiced no strong opposition to education reform. However, there is real opposition to performance assessments from groups opposed to outcomes-based education on principle (because they fear that schools would be teaching values contrary to their own). To quell these groups' anxiety, the school system promised that education would not explicitly focus on teaching *values*, and would measure only academic outcomes.

School Context

Ann Chester is an unusual school in many respects. It opened its doors in the fall of 1992, with a mission to establish a "distinctive learning environment" for its students. Beginning in 1993-94, it reorganized into a year-round school with nine-week sessions alternating with three-week breaks. The school's two Teacher Directors share administrative responsibilities with other teachers, parents, and community members through the school-based management (SBM) team.

---

[4] The PACE program is directed by Dennie Palmer. It is based at the Harvard School of Education and is funded by the Rockefeller Foundation. Twenty middle schools in the district are participating in it.

They view themselves as "voices" in the school's decision-making process, not as definitive decision makers.

The school also is mindful of its multilingual student body, and promotes Spanish as a second language. Four certified bilingual teachers and parent volunteers teach ESL children. In addition, the school offers its students Korean language classes on Saturdays. This Saturday school program will expand next year to include Chinese and Spanish.

Ann Chester is, then, a hotbed of reform activity, and the manifestation of assessment reform at the school can be understood only within the larger context of its broad philosophy of education.

Ann Chester's School Mission[5]

Ann Chester's stated mission reads as follows:

> *To establish and maintain a learning community which challenges traditional assumptions about the educational process in a way that:*
>
> - *Creates a child-centered environment where students have choices, make decisions, and accept responsibility;*
> - *Involves students in a meaningful learning experience which includes applied learning projects; and*
> - *Provides for broad-based input by classrooms, community, and corporations in the decision-making process, so that students acquire a foundation of real world experiences to become motivated, successful, lifelong learners and community participants.*

The school's official brochure states that parents, school staff, students, and community members will collectively develop and maintain standards of student performance. It states that at special forums, this group will decide:

- *What is important for students to be able to do;*
- *What knowledge students should be able to demonstrate; and*
- *What standards are considered appropriate.*

---

[5] From an undated school document.

Student progress with regard to these criteria is to be assessed (a) by students themselves, (b) by the observations of teachers and parents, and (c) through the maintenance of student portfolios.

Ideas deriving from the *New Standards Project* and the *Applied Learning Project* have helped mold Chester teachers' approach to setting standards and to evaluating their students. Chester teachers plan to refine grade-level performance standards for English language arts, social studies, mathematics, and science.

New Standards Project Performance Assessments

Chester employs the New Standards Literacy dimensions to set 4th-grade standards for reading and writing, the elements of which are: *Constructing Meaning; Craft Consideration; Strategic Use of Conventions; Ownership; and Strategic Use of Processes* (see Appendix A for the NSP literacy dimensions). Chester teachers are now producing a report on Chester students, using the NSP grade 4 rubrics, exemplars, and commentaries from this task. (This report is not yet available.) The school also plans to adopt the NSP mathematics dimensions.

Three teachers at Ann Chester are currently developing and piloting mathematics tasks and associated scoring rubrics, which they then send to the NSP (and for which they receive some money if the task is adopted). The teachers utilize the National Council of Teachers in Mathematics (NCTM) standards for generating their performance tasks. The most difficult aspect of this process, said one teacher, ". . . is developing and refining the scoring rubrics."

School-based Performance Assessment System

The school's culture is being "redefined," which means, essentially that teachers are enjoying more autonomy in the classroom. Hence, while the NSP and the ALP provide ideas and support, an individual teacher's creativity gives shape to the programs implemented within the classroom.

Each teacher employs a variety of non-multiple-choice, extended-response assessments on an ongoing basis. They tend to blend assessments with instruction, so that the mix garners diagnostic, instructional, and curricular information. Teachers also use assessments to provide information to students and to parents about student educational progress. Because of this mix, no clear distinctions exist between assessments and instruction. Assessments also are used for advancing students from one grade to another.

Instructional Approach

Classrooms at Chester are filled with all manner of objects — fish tanks, boxes, pictures, chairs, and sofas. No classroom is arranged in old-fashioned rows, and students have the freedom to move about so that they can use resources and otherwise engage in project work (see Exhibit II). Teachers use a variety of pedagogical approaches, from cooperative learning to individual study-time. To prepare students for the TAAS, however, teachers analyze the TAAS specifications and plan activities for students to practice the TAAS format, which is multiple-choice.

EXHIBIT II

Observation of a 3rd Grade Classroom

Students formed groups of threes or fours and read an assigned mathematics word problem. Some worked on the problem alone, while others wandered around the classroom and looked over other students' shoulders, discussing different problem-solving strategies. Other students also sat on the floor, on chairs, or simply lounged at their desks. The teacher sat at one end of the room and addressed students who came to her for help.

In some respects, the assessment strategy appears to have paid off in terms of the TAAS pass rates. For the 1993-1994 academic year, the school attained an "acceptable" rating on the test. The percentages of students in grades 3, 4, and 5 passing the reading and writing tests were higher than or quite close to those of the state as a whole. In math, however, the percentage of Ann Chester 4th and 5th grade students passing the test was lower than the figures for the state as a whole.

Assessments

Projects. Ann Chester teachers use different types of assessments, all of which are discussed below. All assessments are hands-on project work that is inquiry-based and designed to simulate the world of work (a direct influence of the ALP). "Project work is the heart of our work," said one teacher-director.

In 1994, for example, a team of students designed their own playground. In order to complete the project, they had to (a) understand and manage cost issues, (b) obtain materials, (c) contact builders, and (d) execute the project themselves. Other hands-on projects include news broadcasts by the "news crew" and a student team designed application forms and procedures. All projects explicitly juxtapose academic learning and social problem–solving in order to reinforce the connection between educational and "real life" experiences.

Teacher-student assessments. Teachers are quite confident about the expectations they have for their students and typically do not develop scoring rubrics for the various projects their students conduct. However, they do encourage their students to formulate their own scoring rubrics (see Appendix B).

For example, a 4th-grade teacher guides her students to develop their own standards for performance and scoring rubrics, which she calls *reflections*. Using this method, she attempts to inculcate a sense of responsibility and standards in her students. She then uses the student-developed standards to determine the student's strengths and weaknesses. Some teachers have devised other diagnostic measures, such as keeping written notes of their children's performance on reading, writing, math, and "mile-stone" events.

Even though these individualized assessment methods indicate that much instruction is tailored to the student, many teachers believe that instruction could be further individualized.

Portfolios. Each Chester student is expected to keep a portfolio of his or her accomplishments in reading, writing, and projects that indicate that she or he is progressing as a learner. Prescribed guidelines indicate that the portfolios should include:

- A list of readings done over the year with an explanation of why the student chose those readings, what they liked the most about what they read, and why;
- Different types of writing done over the course of the year, including a favorite piece with an explanation as to what makes the piece special; a piece detailing the strategies the student uses as a writer; and
- Assignments completed over the course of the year that indicate the student is a good learner. Such assignments can include mathematics problems, drawings, or some special project. [From school document]

Each student's portfolio thus contains samples that cut across all content areas; and students routinely revisit these portfolios to assess their progress vis-a-vis pre-defined goals they set at the beginning of the year. The information in these folders is used to generate narrative report cards. Once a year, the school invites parents to an *Assessment Night* to examine their children's portfolios (see Exhibit III).

EXHIBIT III

Observation of an Assessment Night

One Assessment Night was held from 7:00 to 8:15 p.m. on May 25, 1994. The school theater was packed with parents and students, and one teacher-director welcomed the group for the children's "moment of glory." Next, the NSP coordinator invited parents to examine the NSP's literacy standards, without mentioning the NSP. She asked for parental support in adopting the criteria and stressed that the standards would enable comparisons between Fort Worth students and the national student population, not just between the Fort Worth students and the ". . . kids down the block." She emphasized the national character of what she called the "project."

Next, the parents went into the classrooms and examined their children's portfolios. They discussed the portfolios with their children and appeared to be impressed with the various projects. They provided feedbacks to their children and teachers on a form.

In addition to portfolios, schools maintain folders that contain state-mandated documents such as (a) the student's year-end evaluation, (b) test information, (c) language assessment results, (d) special programs information, and (e) personal data, such as age and grade.

Narrative Report Cards

As an extension of project-oriented instruction and assessments, Chester received a waiver from the district allowing it to replace the traditional letter-based report card with a *narrative* report card (see Appendix C). (It is the only school in the district to have obtained such a waiver.) Teachers use the narrative report cards to describe individual student strengths, weaknesses, and overall progress over the course of a term.

Initially, the narrative report included descriptions of the student's *dispositions of character and habits of mind*. However, to assuage parents' concerns regarding content-related progress, teachers included content-related progress assessments as well. Content area words are now highlighted, and a statement about the student's rank in the classroom is included.

Parent-Teacher Conferences

Chester holds parent-teacher conferences twice a year, with the objective of setting goals for each child for the coming months. For example, one child and her parents set a goal for the child to become "more interactive and expressive in the classroom." With her teacher's help, the student made in-class presentations to overcome her shyness.

Parent-Student Conferences

Student portfolios form the vehicle of the end-of-year student-parent discourse on student academic progress. Student achievement is demonstrated through the projects contained in student portfolios. (This year's portfolio menu was designed to accommodate parents' wishes and also to provide greater accountability to the district.)

Resource and Training Support

The professional climate of the school is palpable. Teachers have established a tremendous network of informal support within the school. Each teacher belongs to one of the math, science, or language-arts *cabinets* that convene once every nine weeks to discuss issues relevant to education and to hold professional workshops. These teacher-initiated study groups mull over assessment, instruction, and curriculum issues after school hours.

For example, one of the NSP 4th-grade teachers informs the mathematics cabinet about the latest developments in the NSP mathematics standards and functions as a *peer-coach* in designing in-class mathematics assessments. Another group interested in literacy is currently reading and discussing *Windows into Literacy*. Visitors from Central Park East 1 Elementary school in New York City, too, have infused new ideas into the school by sharing their assessment and instructional approaches. (Teachers fear, however, that insufficient support from the district for instructional and assessment reform might subvert this process of change.)

The culture of innovation in instruction and assessment at Chester derives vigor and direction from research materials as well. One teacher-director was clearly impressed and influenced by Grant Wiggins' and Lauren Resnick's writings. The ARTS PROPEL project from Pittsburgh and Dennie Wolfe's PACE program also have provided valuable information.

Seventeen of the 18 Chester teachers also have participated in the district's Applied Learning training sessions with the program coordinator. Becoming a fully trained Applied Learning teacher entails participating in three summer training sessions and four training sessions during the school year. The Applied Learning coordinator also is instrumental in finding funding for teachers to attend other professional development programs.

The ALP director has given support, as well, to the NSP activities. For example, she arranged substitutes for teachers who wanted to attend the NSP meetings. She also played a central role in helping to generate ideas for designing narratives and portfolios.

Despite the support network available, teachers desire more professional time within regular school hours. For example, teachers can easily see the structure of the day altered to include blocks of time for *merged classes*, requiring one teacher to be in class with two groups of students while allowing the other teacher to use the time for activities such as designing assessments and writing scoring rubrics. (Currently, teachers have 45 minutes of planning time.

Impact of the NSP and Other Performance Assessments

Currently, no formal evaluation studies elucidate the effects of assessment reform at Ann Chester. Yet, teachers and others feel that the ALP and NSP programs have contributed to the success of their students — believing, in short, that those students are making "good educational progress."

The ease with which teachers have adopted performance assessments and project-based instruction illustrates how much is owed to an aggressive leadership. The two teacher-directors are "pro-reform," for both the NSP and the ALP, and believe in the instructional quality that the new assessment approaches induce.

Ann Chester teachers trust their approach to education, and believe that their students will do well on any measure of performance, be it multiple-choice or performance assessments. Hence, it is difficult to determine the impact of assessment reform at Chester.

Teachers believe their system to be quite reliable and valid (for each student) and are confident that they hold their students to high expectations.

For example, one NSP participant teacher said that he explicitly used performance assessments to test his instructional strategies. In this teacher's "math-centered classroom," students conduct projects with a focus on mathematical activities. In addition, all students play chess as a means of sharpening their reasoning abilities.

To test the viability of his instructional approach, the teacher first developed and administered to his class 10 open-ended assessment items covering specific mathematics areas. Next, he assigned several math-centered projects to his students and retested them on the open-ended assessments. He believes his system to be rigorous, as the students exhibited improved performance upon being retested. To evaluate his students' mathematics abilities, he utilizes the NSP criteria, but no defined scoring rubrics.

This teacher's evaluations are completely without number or letter grades, a system which the teacher finds to be "more authentic." He finds that the use of performance assessments has improved his instruction and allowed him to listen more to students' explanations of their answers. He also finds that his students have become quite proficient in mathematics operations as a result of the project-based instruction.

This teacher, however, is cognizant of the importance of the TAAS. Therefore, to train his students for the TAAS, he translates project-based tasks into multiple-choice test items and administers them to his students. He wants students to be able to recognize distractors such as *wrong choices*, on those tests.

Impact on Students

According to teachers, performance assessments and project-oriented work have had unequivocally positive effects on their students. For example, one 4th grade NSP teacher said that his students ". . . bubble all the time. They feel confident."

The students interviewed for this study had similar reactions. For example, one student described projects as "fun activities." "My portfolio has my best work and some of my favorite work," declared an articulate 5th grader.

This student also likes the fact that she does not have to sit in a row and raise her hand and wait to be acknowledged. She attended another school for first grade where students ". . . had to raise hands and sit in rows, read textbooks. It was boring." She likes being able to write her own stories now. However, another student admitted to getting tired of the amount of writing required of him. "Sometimes I get tired of writing all the rationales, but it's better than not writing," said this 5th grader.

Students are familiar with scoring rubrics and construct their own in order to evaluate their work as well as that of others'. Students said they like being able to brainstorm and discuss what constitutes an *excellent* performance.

Chester teachers are impressed with Chester students' vocabulary, ability to reason, willingness to challenge their own and others' thinking, and ability to provide rationales for their opinions. Teachers also believe that many students have matured socially, take criticism well, and are able to work cooperatively in groups.

Teachers surmise that the main reason students are able to develop and exhibit these abilities is because they are engaged in solving social problems that truly motivate them to do well. In addition, project-based work enhances student understanding of subject matter as it accommodates different learning styles, enabling students to learn and understand in their own ways.

Teachers added that they have observed no gender or ethnic differences in student performance.

Special education students. Teachers, however, evinced mixed reactions about special education students performance on project-based work and assessments. Some believe that these students are performing well while others feel that they are struggling with project-oriented work. The effects of Chester's project-based pedagogical approach on special education students is unclear.

Impact on Parents and the Community

Parents choose to enroll their children at Chester. It is therefore not surprising that parent participants of this study are generally satisfied with the school and with their children's progress. They like the active-learning pedagogical approach and the narrative report cards, which help them to learn about their child's strengths and weaknesses.

However, one parent questioned whether the narratives are actually tailored to reflect individual students' educational attainments and knowledge. He said, that as an evaluator for Lockheed, he was "formulating similar types of evaluations" and "could only produce so much variation in them." Furthermore, he remarked that the primary question on parents' minds when they see a narrative report card is: "Is Jimmy number one?"

Parents also simply enjoy leafing through their children's portfolios at the end of the year. A 2nd-grade parent remarked, "Portfolios allow kids to illustrate progress and work [and] give parents deeper insight into kids' learning." This parent felt that compared to his older daughter, who went to a traditional school, his 2nd-grade Chester child is learning "much more."

Despite such accolades, parents did express concern about their children not gaining enough exposure (drill and practice) to what, in their estimation, are "basic skills." They want their children to know the multiplication tables as well as spelling and grammar — and they want their children to be able to perform well on traditional tests. As a result of these desires and concerns, one teacher allows parents to take children out of her class to drill them in multiplication tables and spelling. In addition, the school organizes forums for parents to inform them about how basic skills are being taught through the Applied Learning projects and performance tasks.

The community. The board member who participated in this study is supportive of Chesters educational approach. He believes that students should be evaluated "against standards, not against one another." According to him, other board members, too, are largely supportive of Chester's philosophy and operations. They believe that ". . . the best way to reach disadvantaged kids is through active-learning" and education reforms are most effective when they focus on the basic process of instructional delivery.

Nonetheless, there is some tension in the community, particularly around issues of content coverage. "When you do applied projects, you can't sacrifice content," said the Board member. He feels that the "principal focus has been process," and believes that traditional content areas must be "infused" into the curriculum.

Future Plans

The district also hopes to compare student performance on the NSP tasks with performance on the TAAS tasks, in one or two years. NSP itself will conduct an evaluation of its system in two to three years.

Ann Chester teachers will continue to use the NCTM and NSP standards to define the school's goals and standards for student achievement.

Conclusions

The FWISD has focused more on investing in and building teacher capacity through instructional techniques and assessment training programs than on refining a particular pedagogical approach or assessment system per se. Thus, the presence of the NSP in the school is shaped through its *Applied Learning* philosophy. While only three Chester teachers are directly involved in the NSP program, many more are participating in the ALP. Thus, it is the convergence of useful ideas from the two programs that is transforming education at Chester.

Chester typifies the philosophy of intertwining assessments and instruction as teachers see fit. The school values individual goals and progress, prescribing only general assessment and instructional formats, many of which are drawn from the two aforementioned programs. As a result, teachers use a variety of approaches toward assessments, including children's assessments of themselves, as pedagogical tools.

While the *process of reform* appears to be smooth at Chester, whether it is efficacious with respect to student achievement remains to be seen. The outstanding and controversial issues that remain to be addressed at Chester center on (a) standards, (b) content coverage, and (c) format for reporting student progress. The school is forced to balance the innovative with the traditional. Finally, though, Chester teachers will continue to focus on developing performance standards for each grade-level in each subject area, and they will continue to collect student work that demonstrates "how good is good enough" relative to their standards. This dedication to reform will lead, they hope, to a student achievement level with which everyone is satisfied.

APPENDIX A

NSP Literacy Dimensions

DRAFT

Dimensions (rev. 5/21/94)

	Constructing Meaning	Craft Considerations	Strategic Use of Conventions	Ownership	Strategic Use of Processes
Reading	•summarizes •gains information •analyzes, interprets, evaluates text •personal response to text, i.e., connecting to experiences or other texts	•recognizes author's craft and style, i.e., literary styles, language, etc. •recognizes a variety of genres •recognizes functions of print conventions, i.e., boldface, italices, formatting, presentation, type size, etc.	•fluency and self-correction •demonstrates ability to access a variety of resources for a variety of purposes, i.e., to gain information, read for pleasure, etc.	•self-reflects and evaluates self as a reader •participates in a community of readers •reads widely and for a variety of reasons •self-directed choice in reading materials	•predicts, questions, re-reads, decodes, seeks information, consults dictionaries, consults other members of their literacy community, uses reference materials to solve problems, uses context to construct meaning •
Writing	•presents ideas clearly and effectively •develops ideas with appropriate information or supporting detail	•writes for a variety of purposes and audiences •demonstrates effectigve organization, i.e., easy to follow, transitions, introductions, sense of closure •uses precise and varied language suitable to topic •writes sentences with a natural flow of language	•uses appropriate paragraphing, capitalization, punctuation, spelling, and usage (fix this)	•uses writing to think and learn •reflects on and evalutes on growth and accomplishment as a writer •demonstates independence and persistence •participates in a community of writers	•uses a writing process, i.e., planning and prewriting, drafting, revising, and editing
Language	•presents ideas clearly and effectively •supports main ideas with relevant detail •demonstrates control of ideas and development •participates in discussion •	•considers audience with respect to purpose, format, content, and delivery •is easy to follow, i.e., effective organization, transitions, introductions, sense of closure, etc. •monitors, evaluates, and adjusts	•uses standard English when appropriate •varies use of language according to audience and cultural and community contexts	•builds upon home language •reflects and evaluates self as speaker •presents ideas confidently •takes risks, i.e., tries new ideas, takes a controversial position, works in unfamiliar ways	•speaks fluently, i.e., minimal use of space fillers •uses a process to organize, develop and deliver an oral presentation,i.e, gathers information from multiple sources, plans, and uses audiovisual aids

BEST COPY AVAILABLE

APPENDIX B

Reading Log Rubric

Reading Log Rubric

FOR CHILDREN GRADE 3

Poor	Okay	Outstanding
Missing date, author, title, or chapter	Date, author, title, and chapter included	Date, author, title and chapter included
Difficult for others to read	Can be read by others	Others can read it easily
Summary-Little information about the book or chapter	Summary-Tells some information about the book or chapter	Summary-Tells a lot of information about the book or chapter
No reaction to the story or little thought put into the reaction	Includes some reaction to the story	Thoughtful reaction to the book or chapter and may include personal reflection and/or questions

APPENDIX C

Narrative Report Card

NARRATIVE REPORT

NAME First Reporting Period

THE STUDENT AS A LEARNER:

NAME appears to have little or no difficulty directing his own learning. He is resourceful in obtaining various materials and people resources needed to guide his learning. He usually makes appropriate choices in his work and is able to take responsibility for those choices. However, he doesn't always seem to see the importance of meeting deadlines, and he often settles for turning in work that is less than his best. NAME appears to be able to identify and solve problems that affect his learning by using logic and reason.

In working with others, NAME is generally a good listener and is not afraid to contribute. He is basically a leader, but he is also willing to work with others and is able to compromise when the need arises. He appears to have high self-esteem, and is able to accept encouragement as well as other feedback. He is also good at remaining objective and helping other students resolve problems.

UNDERSTANDING THE CONTENT AREAS:

In Language Arts, NAME seems to be very comfortable with reading, and is able to choose appropriate reading materials. Before reading a text, NAME is able to predict what will happen based on the title and pictures. When reading aloud to the teacher NAME exhibits many skills associated with good readers. His reading rate is high and fluent, he uses expression, and he attends to punctuation. He makes very few mistakes, but when he does, he is able to reread to correct or confirm what he has read. He successfully uses sound/letter relationships, word order, sentence structure, and his own background knowledge to figure out difficult words. After reading a passage, NAME is able to retell all significant details in a sequential order. It seems clear that NAME is reading in order to gain meaning.

NAME has experimented with a variety of writing types this first nine weeks of school. His narratives (stories) contain an introduction that sets the context of the story, he follows a logical sequence of events and attempts a conclusion. He appears to be able to apply rules of capitalization in his writing, and he attempts to apply rules of punctuation and to use complete sentences.

BEST COPY AVAILABLE

NARRATIVE REPORT

NAME First Reporting Period

Other types of writing NAME has tried with include summarizations of learning and a manual. He is getting better at summarizing what he has done and learned for the week. His manual had a cover, table of contents, and an introduction. He did a nice job of explaining the newspaper jobs, but he left out a lot of important details in putting the paper together and actually publishing it. It doesn't appear he revised his manual. This is an important part of the writing process.

In Math, and Science, NAME is generally able to choose appropriate problem solving strategies and explain his work. He is also learning to apply what he has learned to other situations, like selling ads for the school newspaper or figuring stocks. NAME knows some of parts of his multiplication facts, but needs to work on learning the rest of them. In the science center, NAME asks appropriate questions and makes relevant observations. He also exhibits the ability to compare and contrast various items, such as rocks and shells, and he is able to draw logical conclusions from what he observes.

NAME is willing to participate in Social Studies discussions. Through our current events study, he has had an opportunity to compare and contrast the various ideas and beliefs of the different presidential candidates. During class meetings, NAME often volunteers to lead in discussion of compliments and/or problems. This is an important part of a plan to help students get along better. He is also learning to work with others, and how to interpret the rights, duties, and responsibilities of citizens.

Overall, NAME is doing well. He is functioning with other students at the top of his class.

BEST COPY AVAILABLE

APPENDIX D

Parental Feedback Form

End-of-year Portfolio Presentation
5-26-94

Feedback to ______________________________ from your family
(child's name)

I have reviewed your 1993-1994 portfolio and particularly enjoyed seeing these things:

From your portfolio presentation I have evidence that you have grown as a learner this school year. Specifically, these are the artifacts which indicate growth in the areas listed:

artifact/evidence of growth	area of growth
(journal entry, graph, book list, research paper, description of event, project work, experiment, photo, tape, other)	(reading, writing, science, mathematics, problem-solving, self-initiative, communication, other)

I will support your work as a learner during the 1994-1995 school year in these ways:

When you set your goals for the next school year, I recommend that you consider including these two or three goals:

Thank you for sharing your portfolio with me/us

____(family signature/s)

The New Standards Project: Noakes Elementary School February 21-22, 1995

THE NEW STANDARDS PROJECT:
NOAKES ELEMENTARY SCHOOL

Introduction

This case study focuses on the piloting of the New Standards Project (NSP) and the implementation of locally developed performance assessments in Anton, Iowa — a district that began participating in NSP due to its growing interest in alternative forms of student assessment. Anton's District Superintendent viewed the introduction of performance assessments as a step toward establishing an educational system that is more connected to what he calls "making learning real." He also wanted to ensure that Anton students would be able to compete in what he views as a changing world, in which the children of today must be prepared to do well in the future. "Our kids are doing well today, but what about tomorrow?" he said.

Noakes Elementary School enrolls 565 kindergarten through 6th graders. The student body is primarily white (98 percent), with the remainder being Asian (1 percent), Hispanic (2 percent), and African American (1 percent).[1] Four percent of the students are eligible for Chapter 1 funds and 7 percent are eligible for free or reduced price lunches (statewide, 13 percent of Iowa students are eligible for Chapter 1 funds). The school has a 97 percent attendance rate.

Participants

The individuals interviewed for the study are listed in Exhibit I.

EXHIBIT I

Study Participants

- District Superintendent
- District Executive Director of Assessment
- One School Board Member
- State Department of Education Consultant
- Two elementary school Principals Students
- Two High School Teachers
- One 4th grade Teacher
- Two 6th grade Teachers
- One Parent of a 4th grader
- Four High School Students
- Four Elementary School Students

---

[1] Total does not add to 100 due to rounding.

Observations

At Noakes Elementary School, the piloting of the NSP Language Arts assessment task was observed. At Anton High School, an Honors English class was observed during its weekly "portfolio day." A meeting of the District Assessment Committee also was attended.

New Standards Project

Jointly run by the national Center on Education and the Economy and the Learning Research and Development Center at the University of Pittsburgh,[2] the New Standards Project is a partnership of 17 states and 6 school districts working toward developing and adopting both a set of national education standards and new ways to measure student progress. Anton has piloted English language arts (ELA) classes in the 4th, 8th, and 10th grades, and has volunteered to pilot 4th and 5th grade Math performance assessment tasks in the near future.

State Context

Because Iowa is one of two states without a specific state-wide testing requirement, NSP was interested in soliciting the participation of several of the state's districts. Currently, all districts must report yearly progress of student achievement to the State, but are allowed to choose how they measure and show progress. Most districts use the Iowa Tests of Basic Skills (ITBS) for reporting purposes, and have consistently attained high test scores on such national tests.

It is important to note at the outset that, in Iowa, the majority of decisions regarding education are made at the local level. It is up to the districts, to decide whether they should utilize NSP; and although only 5 districts were piloting NSP assessments in 1991, by 1995, 65 Iowa districts were participating in the NSP program. State support for these districts includes $150,000 a year for the state to belong to NSP. According to one NSP teacher, the Legislature believes that NSP has thus far been a "good investment," primarily because of the number of teachers NSP has trained and the amount of materials that have been provided to state schools. However, the state mandates no specific requirements regarding the use of NSP assessment results; each district determines how it will use those results.

District Context

Anton is located in a suburban area of metropolitan Des Moines. As Anton has grown from a small town to a suburb, its story of education has been one of increasing enrollment and continuous expansion. During the 1994-95 school year, for example, the district's enrollment was 4,889, up more than 20 percent from 1984-85. The district currently consists of six elementary schools, one middle school, one high school, and is in the process of constructing a $9.5 million middle school.

---

[2] NSP is discussed in greater detail *The New Standards Project: Ann Chester Elementary School.*

Similar to the figures quoted with regard to Noakes Elementary, the district's student population is 97 percent Caucasian, 1 percent African American, 1 percent Asian, and 1 percent Hispanic. Seven percent of its students qualify for free or reduced price lunches, while only one percent qualify for Chapter 1 funds. Typically, Anton students perform above grade level and rank near the top on national standardized achievement tests. In fact, on the 1993-94 ITBS, Anton students in grades 3-8 were, on average, in the 90th percentile on national rankings; and Anton 10th graders had an average composite ACT score of 23.0, compared to the national average of 20.7. Seventy percent of Anton graduates enroll in some form of postsecondary education.[3]

The Anton school district is committed to involving the community in its current operations and in its future plans, primarily because the district sees its own success as both resulting from and contributing to the overall success of the larger community. In fact, to demonstrate its commitment to the community, the District provides community members the use of the public school facilities for adult education classes and community betterment activities (with nearly 500,000 hours of community use scheduled annually in local school facilities). Because committees are a way the school district can involve community members as well as teachers and administrators in the district's decision-making process Anton has used a "committee approach" to governance sin the mid-1980's. Furthermore, according to the District Superintendent, Anton's School Board is a "working board, not just a rubber stamp." These three factors — Anton's commitment to community service, community participation, and an involved school board — help the district to keep abreast of community opinion to and address any concerns parents and community members might have about the school system.

History of Assessment

Although NSP membership is an important aspect of Anton's push toward performance-based assessments, Anton teachers' involvement with performance assessments actually began over 10 years ago. At that time, the English Language Arts department at Anton High School started experimenting with holistic scoring and rubrics designed to guide students in learning various writing skills, and the district began scoring 5th, 8th, and 11th grade writing holistically. It was, in part, Anton teachers' long exposure to portfolios and performance assessment that convinced them to join the NSP in 1992. They realized that improvements in student achievement are tied not only to the use of performance assessments, but also to staff-development and an improved curriculum, the two objectives NSP emphasizes in its program. Furthermore, several teachers mentioned that one reason Anton is involved in NSP and performance assessment reform in general is that there is a core group of teachers, parents, school board members, and administrators who are willing to try new ideas.

Once committed to encouraging the development of new performance assessments, Anton applied for and received a five year waiver from the state, which releases the district from having to report measures of student progress. Currently, Anton is not using NSP results for reporting purposes. However, the district is considering using NSP results for Chapter 1 funding, but will pilot the idea before any decisions are made.

---

[3] District publication, *Ankeny Community Schools — Excellence in Education (1994-95).*

Anton's formal participation in NSP is through two 4th grade teachers at Noakes Elementary School and, at the high school, through four teachers using NSP portfolios and two teachers who are piloting NSP English language Arts (ELA) tasks. Other teachers in Anton also have started developing their own tasks and sharing them with colleagues and, in some cases, with NSP, for feedback.

Noakes piloted its first NSP 4th grade ELA task in May of 1992. Since then, the school has piloted numerous other ELA and Mathematics tasks and, currently, has a two-year commitment to pilot the NSP ELA portfolios in the 4th grade. The first ELA reference examination will be administered in the Fall of 1995. The ELA exam lasts two days and combines questions similar to those in the performance tasks with multiple-choice questions. The mathematics reference examination already has been developed and the second pilot was administered in April of 1995.

Although not the primary focus of this report, students and staff members at Anton High School also were interviewed. Currently, two 10th grade teachers are using NSP pilot performance tasks with their students and four high school teachers are requiring their students to compile NSP ELA portfolios — with student opinion solicited prior to adopting the portfolio system. Students themselves debated the characteristics and benefits and drawbacks of portfolios and then voted to use them, demonstrating Anton's commitment to total involvement.

Assessment Characteristics

NSP Performance Tasks

The NSP English Language Arts performance tasks are open-ended, requiring extended written responses. These tasks were piloted in the 4th and 10th grades, during individual class periods, and typically lasted for 3-to-5 days. In a standard 4th grade NSP performance task (see Exhibit II for a complete description), students are given two booklets. The first booklet contains the story the students read for the task. The second booklet is the response booklet, which contains all of the instructions, questions, and prompts the students need for the 4-day assessment. Students write all of their answers to each activity in the second booklet, write and revise the final essay in the booklet, and present their final draft. During the assessment, students are allowed to make revisions and have other students review their work. Assessments similar to NSP tasks are used by Anton teachers in individual classrooms.

Scoring rubrics. Each NSP ELA performance task has its own set of anchor papers and scoring rubrics; and teachers receive a generic reading rubric and a specific writing rubric for each task. Teachers trained by NSP said they are comfortable with the reliability of the NSP assessments because the students are shown rubrics and are allowed to discuss expectations "up front." The 4th grade NSP teacher said that she is comfortable scoring tasks because she has been trained. She said, however, that a good, clear set of anchor papers is essential to good scoring. Another teacher mentioned that the tasks gave him ". . . a clear picture of what students took away and whether they met [his] objectives."

EXHIBIT II

Piloting NSP Camels Task

Before the task began, the teacher provided her fourth grade English Language Arts class with an overview of the "Camels" task — what it was about, how it would be administered, and what they should think about or concentrate on during the task. During the assessment, the teacher continued to give instructions, prompts, advice, and posed questions for the students to consider as they wrote. The task was administered in 4 days during 45 minute class periods.

Day 1 of the "Camels" task: Students read *Ships of the Desert,* a short story about the history, temperament, and physical characteristics of different types of camels. The instructions told them to think about ". . . whether or not camels are good or bad animals to have around." After the students finished reading, they opened their workbooks and began answering the questions.

Students were asked to answer questions such as "Why might the author have written this article?" and "Do you think camels are good animals to work with? Use at least one example from the article to explain your answer."

Day 2: Students were told they will be asked to select and write about one of two topics: (1) describe an imaginary animal; or (2) recommend whether or not a camel should be added to your community zoo. Students were asked to make lists such as "How camels are useful" and "Why camels are unpleasant," to help them make their decisions. The students then began their first draft (students were not given a length requirement).

Day 3: Students spent 15 minutes looking back at their draft and were given several questions to think about while they read. Next, students got into groups and shared what they had written. Students were given response forms and asked to share their ideas with the group.

Day 4: Students spent the entire period revising their draft.

NSP Portfolios

An NSP portfolio is a collection of work that includes specific samples of student work that demonstrates the ability to communicate clearly and effectively. Portfolios are completed by both Elementary (4th grade) and High School (10th grade) students. To help lead them through this process, students use NSP Student Portfolio Handbooks (which they are given at the beginning of the year). Both the High School and the Elementary School handbooks (which are currently in field trial versions) accomplish the following:

- Introduce NSP;

- Explain the standards students must meet in three major areas of English and language arts (reading, writing, and speaking and listening); and
- Inform students about how to organize and assemble their portfolio (e.g., how to make a table of contents and what should be in their "letter of introduction.")

Elementary ELA portfolios. English Language Arts portfolios at the Elementary level must contain:

- a table of contents,
- a reflection piece,
- a response to literature,
- a student's best pieces that tell a story, demonstrate the collection and reporting of information, demonstrate writing instructions or directions, demonstrate growth, demonstrate the student's best efforts in reading, and
- four free choices (at least one must be a speaking and listening entry).

For each item, the student must fill out an entry slip that explains what kind of work the piece is, who chose it, and why it was chosen. In some schools, teachers involve parents by sending home a letter that asks about their child's portfolio. Questions include: *What area impressed you? What growth did you notice? Do you have any comments or suggestions?*

Anton teachers also are discussing the possibility of developing a computer portfolio that would stay with a student during his or her entire school career, enabling teachers to track student progress and curriculums over a number of years.

High school ELA portfolios. At the High School level, each portfolio must contain the following items:

- a table of contents,
- a letter of introduction,
- an entry slip (see Appendix A),
- six or seven pieces of writing from several categories, such as persuasive, informative, or analytical,
- evidence of reading, such as lists of titles, annotated bibliographies and written responses to literature,
- evidence of listening and speaking, and

- two other pieces that represent the student's best work.

Portfolios can contain anything from "personal writing" (in which the student writes about whatever he or she is interested in) to in-class assignments the student may find particularly interesting or challenging or good examples of his or her ability to explore or analyze a topic. Multi-media presentations using audio, video, and computer graphics also may be included. For each entry, students must fill out an entry form that gives the student a chance to think about what the work demonstrates and about his or her development as a learner.

In the high school 12th grade Honors English class, every Friday is "portfolio day." Students spend the entire English period working on their portfolios. Students can write a new piece for their portfolio, conduct a teacher or peer conference (a requirement developed by the teacher – see Appendix B), write a letter of introduction for the portfolio, fill out the entry slip for each piece, or anything else related to their portfolios. Students often check-out the English Department lap-top computers to work on their assignments in class.

At the end of the year, after the portfolios are completed, students must write two letters. The first letter is to their teacher; it states what grade they believe they should receive and why. The second letter is to their parents, explaining what they have learned and what their portfolio "shows" about them.

Scoring rubrics. Portfolios are scored using rubrics that show the student and the reviewer what characteristics must be present in student work at different grade points for the student to receive the different scores. The high school English portfolio scoring rubric outlines standards for reading and writing and provides guidelines for assigning scores ranging from 1 to 4. (See Appendix C.) The rubrics provide guidelines along four dimensions: *range* and *versatility*, *reflective analysis*, *technical control*, and use of *literacy processes*. For example, a student's work must provide ". . . evidence of reading a variety of kinds of texts: a broad range of genres, styles, cultures, topics is represented" for the portfolio to receive a score of 4 in the Range and Versatility dimension. If the portfolio ". . . provides little, if any, evidence of reading; [or if] what reading is evident represents a very narrow range of genres, styles, cultures, topics, and periods," the portfolio receives a 1 in the Range and Versatility dimension. This rubric was developed by the honors English teacher, who modified an existing NSP rubric.

One teacher said she is struggling with how to score portfolios and how to use those scores, her concern being that parents are "very grade oriented." Even if parents are supportive of portfolios, they still want to see the letter grades they are used to seeing. Therefore, even though the teacher finds it somewhat difficult to score portfolios (because the NSP rubrics are not finalized yet), she will give her students portfolio grades during the final two quarters. At the end of the school year, she will randomly select 10 portfolios to send to NSP for a summer benchmarking conference.

Professional Development and Financial Support

Anton's involvement with NSP professional development is largely through NSP national conferences. Each year, NSP conducts conferences for teachers from participating organizations. Since joining NSP, Anton teachers have attended six different national conferences. For example, in June of 1993, two Anton teachers attended a national conference in Portsmouth, New Hampshire, to benchmark performance tasks in reading: one teacher stayed an extra two days to gain experience in writing commentaries for student work. Other such national conferences have focused on scoring (June 1992, Big Sky, Montana; July 1993, Snowbird, Utah), portfolios (July 1994, Indian Wells, California), and developing new performance tasks (National Summer Conference, July 1992, Phoenix, Arizona).

For the most part, teachers' expenses at conferences are paid for by the district, but teachers do not receive additional compensation to attend the conferences. The participant teachers are then charged with the task of sharing the information from conferences with the rest of the teachers in their school and district. The teachers who attend the national conferences also act as trainers, sharing information on scoring, portfolios, benchmarks, and task development; this is done by conducting in-service presentations throughout the state and by facilitating state-level conferences. Two of the Anton teachers have even designed an informal portfolio training session for the school board.

Anton does not have a required number of staff-development hours. However, every other Wednesday, students are released early, and teachers are required to attend staff development activities. Staff development activities can be building-level or district-wide and can include anything from portfolio training to presentations on child abuse.

Anton's Executive Director of Instruction hopes to establish a professional development center that will house information on student-centered learning and performance-based assessments. She realizes that "there is never enough staff development time, but feels that her work in support of performance-based assessments is strongly supported by the administration; and Anton's Superintendent shares the belief that "teachers should be life-long learners."

According to district officials, in 1994-95, the Anton School District spent $35,000 on the development of assessments, over $5,000 on the preparation or purchase of assessment materials, and $3,500 on assessment scoring. The state provided $150,000 to belong to NSP; a $5,000 Technical Prep grant; a three year, $5,000 Exemplary Projects grant; and additional funds from a Phase III grant. Phase III is a state-funded program that provides money to schools to support the development and implementation of innovative ideas and programs. The Phase III funds have supported summer workshops with outside expert speakers and provided limited compensation for teachers who are involved in the project.

The district spent a total of $970 per teacher on staff development. Professional development is, in large part, funded by a 10-year instructional support levy that now is in its fourth year. This levy, which supports professional development and four other components, provides $150,000 per year for staff development.

Relation to Other District-Level Assessments and Reforms

NSP is considered to be an integral part of an overall reform effort to improve Anton's curriculum and assessment methods. A core group of K-12 Anton teachers are restructuring the curriculum to model cooperative learning, the NSP standards and assessments, Master Teaching, and other district innovations — to bring them all together under one umbrella.

The district is not planning on replacing ITBS completely, however, but has reduced its use. According to one principal, ITBS helps present a clear, whole picture of a student's abilities, and parents would not be comfortable without some type of standardized test. However, ITBS is not used to evaluate teachers; and Anton does not publish scores by school, and the importance of results are being downplayed. According to Noakes's principal, the end of all of this, it is hoped, will be a balance of different types of assessments.

There also is a big push in Anton to infuse technology into the curriculum. Currently, for example, one 6th grade English teacher has his students produce a television news show every three weeks. His students must learn to run the cameras, compile the computer graphics and background music, and write the news stories. This is just the beginning of what may become a computer-literacy push.

Already, both district administrators and teachers realize that teachers must be computer literate to effectively use technology in the classroom. Therefore, all teachers are required to pass a performance-based computer literacy assessment. Should a teacher fail the assessment, he or she is provided support until the assessment is passed.

School Context

Noakes Elementary is committed to providing a supportive environment directed toward educational excellence and the development of each student's potential. Noakes teachers and administrators believe:

- Each child can and will learn;
- The instructional program will be character building in its nature, caring in its treatment, and demanding and professional in its methods;
- A positive self-concept and self-discipline are an integral and necessary part of a child's learning;
- Each student will become an informed, competent, sensitive, and responsible community member; and
- Each individual will be respected (and provisions made for their individuality).

Noakes's Performance Assessments

All Noakes teachers are committed to the use of performance assessments — albeit in different ways. One of Noakes's 1994-95 school-year goals was to ensure that each child complete a portfolio in at least one academic area. All teachers have received training in the development and use of performance assessments, have participated in in-services conducted by staff trained by NSP, or have assessment-related professional materials distributed by the district. Two 4th grade teachers at Noakes have had extensive training through the NSP in the development and scoring of performance assessments in ELA. They also have had training in compiling and scoring NSP portfolios.

Noakes elementary teachers are committed, as a school, to the use of portfolios. The school, however, does not use a standard, school-wide portfolio system; and, within that context, teachers are not required to use NSP portfolio guidelines, but are allowed to exercise individual control over how they want to use the portfolios. As a result of such flexibility, some teachers are using portfolios in only one subject, while others are using them in all subjects. In addition, some teachers do not grade the portfolios, while others incorporate portfolio scores into a student's final grade. Noakes portfolios, thus, are used primarily for pedagogical purposes, not for accountability purposes.

Impact of and Reactions to NSP and Other Performance Assessments

The following section examines the impact of NSP and locally developed performance assessments on the Noakes community.

Impact on and Reactions of Teachers

The teachers directly involved with NSP, either through performance tasks or through portfolios, expressed positive opinions regarding the impact the program has had on teaching. Teachers' commitment to NSP is evident by the amount of time they invest in the program — going to summer conferences, giving in-service presentations, and working the extra time needed to score performance assessments. Teachers take the extra time because they believe that performance assessments help students to develop a deeper understanding of subject matter. One 4th grade NSP teacher said that she was initially skeptical, but was sold when she saw how enthusiastic her students were about the types of tasks they were involved with.

NSP also has had an impact on teachers not directly involved with the program. For example, Noakes teachers are using some of the NSP ideas in developing their own assessments and portfolios, and teachers also are developing their own performance tasks and scoring rubrics with others. For example, at the District assessment meeting, teachers bring some of their tasks to share with the group and to get feedback on how their tasks could be improved.

Overall, teachers — both those directly involved with NSP and those devising and using performance-based assessment — had positive opinions regarding alternative assessments. For example, a 6th grade science teacher who administered a 2-hour performance task on electrical circuits said that although it took more time to complete than a traditional quiz covering the same topic, ". . . it was [a] worthwhile, valuable learning experience that the students enjoyed."

Teachers also had concerns, of course — with a lack of time and an inadequate amount of training and models heading the list. As one teacher said, "It [performance based assessment] puts stress on our free time, after school time, and weekends." Performance-based problem-solving activities take longer to implement and in addition, taches must spend extra time devising scoring and some of the science tasks published by textbook publishers (e.g., McGraw-Hill). An individual teacher also must define learning criteria and objectives.

Teachers said that longer or more flexible blocks of time would make performance assessments easier to implement. In addition, teachers also repeatedly mentioned the need for more planning time. As one teacher put it, "We need all the time we can get . . . and more."

Impact on Instruction and Curriculum

One NSP teacher said that NSP opened new doorways for professional growth by showing that new units could be added to the curriculum and that assessments could provide teachers with a "better picture" of their students. This teacher also said that the NSP ideas have changed the way she, herself, teaches. She believes that the new assessment techniques have not only made her more creative, but also challenge her students more.

Teachers had differing opinions about the use and effects of performance-based assessments on students with differing levels of ability or academic achievement. One teacher mentioned that heterogeneous grouping is a problem, because students with different levels of ability need different amounts of time to read the task. For example, students with a lower level of ability are sometimes unable to complete all the steps in a task. She went on to say that it is important for teachers to adjust their expectations to enable students to work at their own level and pace, if all students are to experience success.

Several other teachers, on the other hand, pointed out that performance-based assessments allow students who traditionally experience academic difficulties to perform well, while simultaneously challenging more advanced students. She argued that upper level students will seek out challenges and work to exceed minimum requirements.

Impact on and Reactions of Students

The 4th grade students interviewed had very definite opinions about the performance task they were currently working on and about performance assessments in general. They found the performance task to be very interesting and engaging, and they enjoyed being able to write about their opinions. They also liked the assessments because they are learning experiences. The 4th grade teacher said her students were in fact "sad" when the assessments were over. These students, however, were divided on the idea of working in groups. Some students thought it was a good idea and that they accomplished most tasks, while others thought that their classmates talked too much in groups and that "you can get the wrong answers."

Teachers point out that performance assessments and portfolios help students to demonstrate what they know, reflect on their progress, and take ownership of their work. One teacher claimed that because performance-based assignments enable her students to gain a deeper understanding of the subject matter, they retain information longer. The teacher did, however, point out that students need a "basic knowledge" of the subject area for this to happen.

A high school Honors English student said that he thought portfolios highlight students' strengths and weaknesses and, therefore, are a much better way to evaluate a student. In his opinion, portfolios add a new structure to the class, adding new dimensions to the traditional aspects of English (e.g., reading Poe and Whitman, writing short stories and poetry, and studying

Greek tragedy). Another high school student said "If I had known we were going to do portfolios all year, I wouldn't have planned to graduate early."

Impact on and Reactions of Parents and Community Members

The district is trying to provide information to parents and community members on changes in assessments, as most parents are unaware of Anton's involvement with NSP. One elementary school principal said, however, that parents are just happy to see their children bringing work home for them to see.

According to a parent of a 4th grader, these parents who are aware of the changes are supportive of the changes. This parent said that she has become more familiar with NSP in the past year. In the Fall of 1994, the school had a pot-luck dinner at which the 4th grade teachers spoke about NSP. Later in the year, at another meeting, NSP teachers explained the relationship between the class curriculum and NSP to the parents. This parent said that she understands the system and has seen its positive effects on her child. She has been able to see her child's progress by looking at his portfolio and is impressed by his ability to communicate his feelings about his work. She believes that "NSP has developed this sophistication in him." She also said that NSP portfolios have helped her son to develop diligence in his study habits, has prompted him to be more responsible, and has helped him to become more creative. She is not sure if he is learning more about the subjects, but she is sure he is learning "broader information that has more real life applications." According to this parent, other 4th grade parents share her excitement for NSP.

The parent interviewed for this study said she has become more involved with her son's school work — and has more discussions with him about writing styles and so on. In parent-teacher conferences, the child is actively involved in sharing his portfolio with his mother — and is able to describe the scoring criteria and explain why he included a certain piece, even though he received a "C" on it. But, the parent does not believe that NSP portfolios or performance tasks force parents to be more involved. "Parents who want to be involved in their child's education will be. NSP won't change those who don't want to be involved," she said. A teacher expressed a similar belief when she said that one of the reasons Anton has been so successful in implementing performance-based assessments is that parents like and want to see change and initiative.

Future Plans

Anton School District does not want to *mandate* the use of performance assessments and portfolios. Those involved want the adoption of the assessments to be a bottom-up decision, and the district officials are comfortable with this. As a result, teachers who are using performance assessments, including NSP, are trying to get more teachers involved by providing in-service training and help for interested schools. In fact, in the 1995-96 school year, the district hopes to establish an NSP math team of three teachers at the high school.

In addition to its on-going work in assessment, the district also has received a state grant to establish an interdisciplinary vocational education program, a component of which will be performance-based assessments. At the same time, district leaders are trying to bring cooperative learning, master teachers, and other related reforms together in a cohesive reform strategy, with project and performance-based assignments and assessments as a major aspect of the reform efforts.

Conclusions

In Iowa, the decision whether or not to use performance-based assessments is made at the district level. The school district in Anton, Iowa, chose to use performance-based assessments and joined NSP in 1992, to help implement performance-based assessments such as portfolios and performance tasks. Each school in the district is allowed to determine how and to what extent they want to use NSP or other types of performance-based assessments. To a great extent, each teacher decides how or even if he or she wants to use performance-based assessments. This bottom-up decision-making process has generated an environment in Anton wherein teachers are not afraid to try new techniques, parents are supportive of the school district's efforts, the administration (at the state, district, and school levels) is supportive of innovation and change, and students are exposed to exciting teaching and assessing methods.

Two 4th grade teachers at Noakes Elementary school, the focus of this case study, are currently piloting NSP performance tasks in both English Language Arts and Mathematics. All of the teachers at Noakes have agreed to use portfolios to help develop a complete picture of their students' abilities. At the high school, two 10th grade teachers are piloting NSP performance tasks and several teachers are using portfolios. Parents and students at all grade levels have responded positively to the portfolios and performance tasks and teachers are reinvigorated to be more creative in their teaching.

Anton hopes to continue the use of NSP and other forms of performance-based assessments and integrate their use with other reforms — such as integrated learning and technology-based learning. The teachers have support from the community and the district and the district has support from the state — so the pieces are in place for the reforms to continue.

APPENDIX A

NSP Portfolio Entry Slip

NEW STANDARDS PORTFOLIO
ENTRY SLIP

Date work was completed ______________________________

Name of student______________________________________

School's Name_______________________________________

Teacher's Name______________________________________

Page Number of this portfolio piece________________
(All the pages in your portfolio must be numbered.)

Type of portfolio entry_____________________________
(Tell which of the required contents this piece represents. For example, if this slip were stapled to the back of your reading log, you would write "evidence of reading" in this blank.)

Standards represented in this work
(List all standards embodied in this work. For example, if you submit multiple drafts of an original story, you would note that the piece met *Writing Standard 1: Technical Competence and Effectiveness*, because your story would be free of spelling, punctuation, and grammatical errors; you would also note that the piece met *Writing Standard 2: Range*, because by writing a story in addition to other pieces, you've shown that you can write in a variety of forms.)

____

____

____

In the space provided below construct a paragraph which includes the following thoughts:

- the reason that you chose this piece
- reflections on yourself as a reader/writer/learner
- connections from this piece to other material you have read, viewed, or heard
- any additional thoughts that this piece brings to mind

APPENDIX B

Portfolio Conference Sheet

PORTFOLIO CONFERENCE SHEET

Date of Conference: ______________________________

Conference Participants:
Writer: ______________________________

Peer or Teacher: ______________________________

Use the space below to summarize the comments made by the writer and the reader during the conference.

Major Suggestions for Change and/or Improvement:

1. ____
2. ____
3. ____
4. ____
5. ____

Signature of writer: ______________________________

Signature of peer or teacher: ______________________________

APPENDIX C

English Portfolio Rubric

HONORS ENGLISH
PORTFOLIO SCORING RUBRIC

READING	4	3	2	1
RANGE AND VERSATILITY	• Provides evidence of reading a variety of kinds of texts: a broad range of genres, styles, cultures, topics is represented. • Responds to a wide variety of kinds of texts in a thoughtful and complex manner.	• Provides evidence of reading a variety of kinds of texts; a range of genres, styles, cultures, topics, and periods is represented, although one of these may be represented more completely than the others • Responds to a variety of kinds of texts; some text responses may be more thoughtful and complex than others	•Provides evidence of reading a limited variety of kinds of texts, although there is some evidence of range in genre, style, culture, topic and period •Responds to a limited variety of kinds of texts	• Provides little, if any, evidence of reading; what reading is evident represents a very narrow range of genres, styles, cultures, topics, and periods
REFLECTIVE ANALYSIS	• Provides substantial evidence that the student makes connections in order to reflect on ideas and to analyze and interpret across texts, disciplines, and genres, among language processes, and between experiences in and out of school • Provides thoughtful explanations for reading choices; uses specific references to personal standards and portfolio purposes	• Provides evidence that the student makes some connections in order to reflect on ideas and to analyze and interpret across texts, disciplines, and genres, among language processes, and between experiences in and out of school; connections that are present may be obvious rather than insightful • Provides clear explanations for reading choices; uses references to personal standards and portfolio purposes	•Provides little evidence that the student can make connections in order to reflect on ideas or to analyze and interpret; the connections that are present tend to be very simplistic or superficial • Provides some general explanations for reading choices; makes little or no reference to personal standards or portfolio purposes	• Provides little or no evidence that the student can make connections in order to reflect on ideas or to analyze and interpret; the connections that are present tend to be vague, general, or unrelated to the student's work • Provides few, if any, reasons for reading choices; contains no evidence of awareness of personal standards or portfolio purposes

BEST COPY AVAILABLE

TECHNICAL CONTROL	•Provides substantial evidence that the student can engage challenging texts, both print and nonprint, to construct meaning and to apply interpretations or acquired knowledge to new contexts • Provides a range of substantial evidence of connections between and among disciplines, texts, genres, and experiences that build on prior knowledge to construct new knowledge • Provides substantial evidence of multiple levels of interaction with a variety of texts; interrogates texts and challenges authors' perspectives and purposes	• Provides evidence that the student can engage texts, both print and nonprint, to construct meaning; the student does not always apply interpretations or acquired knowledge to new contexts • Provides evidence of connections between and among disciplines, texts, genres and/or experiences that build on prior knowledge to construct new knowledge • Provides evidence of interaction with a variety of texts	•Provides evidence that the student has some ability to construct meaning from tests, either print or nonprint; however, the student does not apply interpretations or acquired knowledge to new contexts • Provides limited evidence of connections, building on prior knowledge; connections sometimes, not always, lead to constructing new knowledge • Provides limited evidence of text interactions; interactions may be superficial or with a limited variety of texts	• Provides little, or no, evidence that the student can construct meaning from texts, either print or nonprint • Provides little, or no, evidence of connections or of new knowledge construction • Provides little, or no, evidence of questioning or interacting with texts
USE OF LITERACY PROCESSES	• Provides substantial evidence that the student can rethink and refine ideas in the process of responding to, interpreting, and analyzing various texts and forms of oral communication • Provides substantial evidence that the student used a variety of planning and organizing techniques to discover and explore ideas within various texts and discourses	• Provides evidence that the student can rethink and refine ideas in the process of responding to, interpreting and analyzing various texts and forms of oral communication • Provides some evidence that the student used planning and organizing techniques to discover and explore ideas in various texts and discourses	• Provides some evidence that the student can respond to and interpret texts and forms of oral communication; however, responses are often of a personal nature • Provides limited evidence that the student used planning and organizing techniques; evidence may be limited to specific kind(s) of texts or discoursed	•Provides little, or no, evidence that the student can respond to or interpret texts and forms of oral communication; what responses are present are limited to personal reactions • Provides little, or no, evidence that the student used planning or organizing techniques

HONORS ENGLISH
PORTFOLIO SCORING RUBRIC

WRITING	4	3	2	1
RANGE AND VERSATILITY	• Provides evidence of awareness of diverse audiences; the writer's attention to public and private audiences matches his/her varied purposes for writing • Demonstrates the ability to communicate for a variety of purposes; there is ample evidence of the ability to use a variety of genres, forms, and topics in written communication • Provides substantial evidence that the student's skillful control of a variety of distinctive voices makes the portfolio richer, more interesting, and more focused • Provides substantial evidence that the student has attempted to create a portrait of him/herself as a learner by experimenting, attempting imaginative or unusual pieces, or approaching a topic or text in an innovative way	• Provides evidence of an awareness of several audiences; there is attention to both private and public audiences, although one may be primary, while the others are secondary • Demonstrates the ability to communicate for several purposes; there is some evidence of the ability to use a variety of genres, forms, and topics in written communication • Provides some evidence that the student's control of a variety of distinctive voices makes the portfolio richer, more interesting, and more focused • Provides evidence that the student has made some attempt to create a portrait of him/herself as a learner by experimenting or approaching a topic or text in an innovative way	• Provides evidence of some awareness of audience, although much of the work is produced for the teacher as evaluator • Demonstrates a limited ability to communicate for a variety of purposes; there is little evidence of the ability to use a variety of genres, forms and topics in written communication • Provides little evidence that the student can control a distinctive voice; the use of voice is, for the most part, limited to personal responses • Provides evidence that the student has made an attempt to create a limited portrait of him/herself as a learner	• Provides limited or no evidence of an awareness of audience; most work is produced for the teacher as evaluator • Demonstrates little, if any, ability to use various genres, forms, and topics in written communication • Provides almost no evidence that the student can use a distinctive voice to enhance his/her writing • Provides evidence that the student has made little or no attempt to create a portrait of him/herself as a learner
REFLECTIVE ANALYSIS	• Provides substantial evidence of the ability to assess the strengths and weaknesses of own work, processes used, and progress; overall, these self-assessments are both accurate and insightful • Provides clear, thoughtful, thorough explanations of decisions about the reasons for including particular portfolio selections; these explanations reflect an understanding of the portfolio's purpose • Provides substantial evidence that the student makes connections in order to reflect on ideas and to analyze and interpret across texts, disciplines, and genres, among language processes, and between experiences in and out of school	• Provides evidence of the ability to assess strengths and weaknesses of own work, processes used, and progress; although these self-assessments are accurate, they may not be particularly insightful • Provides clear explanations of decisions about reasons for including particular portfolio selections; although these explanations sometimes lack thoroughness, nevertheless, they reflect a general understanding of the portfolio's purpose • Provides evidence that the student makes some connections in order to reflect on ideas and to analyze and interpret across texts, disciplines, and genres, among language processes, and between experiences in and out of school; connections that are present may be obvious rather than insightful	•Provides some evidence of the ability to assess own work; these self-assessments tend to lack focus on products and lack the specificity and accuracy necessary to gauge whether the student has the ability to assess his/her own strengths and weaknesses • Provides some explanations about reasons for including particular portfolio selections; because these explanations tend to be general, they reflect little or no understanding of the portfolio's purpose • Provides little evidence that the student can make connections in order to reflect on ideas or to analyze and interpret; the connections that are present tend to be very simplistic or superficial	• Provides little or no evidence of the ability to assess work; what self-assessments are present tend to be sketchy, general, and inaccurate • Provides few, if any, explanations about reasons for including particular portfolio selections; what explanations are present are so sketchy and general that they reflect a lack of understanding of the portfolio's purpose • Provides little or no evidence that the student can make connections in order to reflect on ideas or to analyze and interpret; the connections that are present tend to be vague, general, or unrelated to the student's work

BEST COPY AVAILABLE

TECHNICAL CONTROL	• Provides substantial evidence that the student can effectively organize ideas and use a range of organizational strategies appropriate to audience and purpose in order to create a unified text • Demonstrates clear command of various types of sentence structures • Demonstrates clear command of conventions and mechanics • Provides clear evidence that the student's word choice and use of language is appropriate to audience and purpose, and enhances the overall effectiveness of the portfolio	• Provides evidence that the student can effectively organize ideas and use organizational strategies that are, for the most part, appropriate to audience and purpose • Demonstrates command of various types of sentence structures, although occasional lapses may be present • Demonstrates command of conventions and mechanics; although occasional errors may be present, they do not cause confusion • Provides evidence that the student's word choice and use of language is appropriate to audience and purpose, and is for the most part, effective	• Provides evidence that the student can organize ideas, although s/he may occasionally present ideas in a random or confusing manner. The organizational strategies used are limited, and sometimes inappropriate to audience and purpose • Demonstrates command of a limited variety of sentence structures; some fragments and run-on sentences may be present • Demonstrates a limited command of conventions and mechanics; some errors may be severe, causing confusion • Provides evidence that the student's word choice and use of language is somewhat limited and may, at times, be inappropriate to audience and purpose	• Provides evidence that the student has little or no sense of how to organize ideas; the student frequently presents ideas in a random or confusing manner • Demonstrates little or no evidence of command of conventions and mechanics; there are frequent, severe errors that often cause confusion and impede understanding • Provides evidence that the student's word choice and use of language is very limited and may frequently be inappropriate to audience and purpose
USE OF LITERACY PROCESSES	• Provides substantial evidence that the student used a variety of planning techniques to discover and organize ideas • Provides substantial evidence that the student used a variety of drafting techniques to develop ideas and to clarify language • Provides substantial evidence that the student used the revision process to refocus, reshape, and refine ideas and texts	• Provides evidence that the student used planning techniques to discover and organize ideas • Provides evidence that the student used drafting techniques to develop ideas, and in some instances, to clarify language • Provides evidence that the student used the revision process, although polished products may reflect only minor changes in thinking	• Provides limited evidence that the student used planning techniques to discover and organize ideas • Provides some evidence that the student understood and used drafting techniques • Provides some evidence of the revision process; however, where multiple copies exist, they reflect little, if any, rethinking	• Provides little, or no, evidence that the student used planning techniques to discover and organize ideas • Provides little or no evidence that the student understood or used drafting techniques • Provides little or no evidence of the revision process; where multiple copies exist, they reflect recopying, and there may be occasional errors

BEST COPY AVAILABLE

McGARY

Harrison School District 2's
Performance-Based Curriculum:
McGary Elementary School
April 11-12, 1994

HARRISON SCHOOL DISTRICT 2'S PERFORMANCE-BASED CURRICULUM: McGARY ELEMENTARY SCHOOL

Introduction

This case study focuses on the development and the characteristics of Harrison School District 2's (HSD2) performance-based literacy assessments, and the implementation and the impact of these assessments at McGary Elementary School, Colorado Springs, CO. HSD2 substantially changed the curricula development process it routinely undergoes in five-year cycles. Simply stated, in the past, those involved in the development of curricula in the district did not discuss outcomes and dealt with the issue of assessment only at the end of the curricula development process. Today, curricula are developed in HSD2 with the explicit consideration of outcomes and assessments at the beginning of the process. The curriculum development process itself involves identifying, integrating, and assessing significant student outcomes *for all students* through the adoption of performance assessments; and through that process, HSD2 is currently working to meet the state mandated (House Bill 1313) requirements for standards-based education in all K-12 Colorado schools.

McGary Elementary is one of the HSD2 schools struggling with and adopting a standards-based education system and related performance assessments. During Academic Year (AY) 1993-94, McGary Elementary (K-5) enrolled approximately 470 students whose proportional composition was roughly similar to that of the district as a whole. Approximately 47 percent were Caucasian, 26 percent were African American, 21 percent Hispanic, 5 percent Asian, and 1 percent were Native American. About three-quarters (72 percent) of the student body receive free or reduced-price lunch.

Participants

The roles of the individuals interviewed for this case study are shown in Exhibit 1.

Observations

The study researchers observed two classes at McGary Elementary School — a peer mediation class preparing for an upcoming retreat taught by the school social worker, and a fourth grade class being introduced to a performance task taught by the building's Literacy Coach. Two additional events were observed — professional development session conducted by the district's *Assessment Academy* for Colorado College's students, and a district assessment scoring session.

State Context

HSD2 personnel and students, parents, and community members are affected by two current issues in the development and implementation of Colorado performance assessments. The first is the passage of House Bill 1313 that mandates standards-based education in every district throughout the state by 1998. The Colorado General Assembly established The Colorado Commission for Achievement in Education, which appointed a Task Force on Student Standards

EXHIBIT I

Study Participants

- Director of Elementary Education
- Principal, McGary Elementary School (works half-time in assessment for HSD2)
- Social Worker, McGary Elementary School
- Four teachers:
  - Fourth grade teacher and Literacy Coach, McGary Elementary School
  - Second grade teacher, McGary Elementary School
  - Fifth grade teacher, McGary Elementary School
  - Eighth grade teacher, Bosse Middle School
- Fifth-grade student, McGary Elementary School
- Parent of a McGary Elementary School student
- School Board member

and Assessment. This task force was charged with the preparation of a plan for standards-based education in K-12 schools. The plan requires that Local Education Agencies (LEA) adopt local standards for reading, writing, mathematics, science, history, and geography — which are to guide curriculum, instruction, material selection, and student performance assessment. (The state itself is developing content standards.) Reading and writing assessments were mandated to be in place by academic year (AY) 1994-95; Math and science assessments were to be in place by AY 1995-96; Diplomas certifying competence were to be awarded after July 1996; and history and geography assessments were to be in place by AY 1997-98. (The state itself plans to develop a new, performance based assessment program for grades 4, 8, and 10, beginning with AY 1995-96.[1])

Secondly, HSD2 personnel's cautious approach toward the adoption of performance assessments was affected by the ouster of the majority of members of the nearby Littleton School Board of Education. Littleton's community members reacted to what they perceived to the be Board's rapid movement toward adopting outcomes-based education (with a performance assessment component). The primary lesson that HSD2 officials expressed they learned from the ouster of the Littleton school board is: If the move to performance assessments is too quick, teachers, parents, and community members will reject the reform. Therefore, as one teacher said, "It is important that we go slow, because of what happened in Littleton."

District Context

Within sight of Pike's Peak, HSD2 sprawls over 18 square miles of Rocky Mountain foothills. The area is the home to over 52 different national and international religious organizations of all types. The United States Air Force Academy and Fort Carson Army Base also call the area home. Over one-third of HSD2 students have parents who are employed by the Federal government, and the transience rate of the district is a staggering 40 percent each year.

---

[1] *State Student Assessment Programs Database, 1993-1994.* The Council of Chief State School Officers and the North Central Regional Education Laboratory.

There are eight school districts in Colorado Springs alone, with every one growing. Over 10,300 students were enrolled in HSD2 in AY 1993-94, representing a 47 percent increase since 1980. As of the spring of 1994, HSD2 operated 2 high schools (grades 9-12), 3 middle schools (grades 6-8), and 12 elementary schools (grades K-5). Nearly 45 percent of HSD2 students are racial or ethnic minorities — the vast majority of whom are either African-American or Hispanic. Nearly 70 percent of the district's students qualify for free (56.4 percent) or reduced-price (12.7 percent) lunch, and approximately 12 percent receive special education services.

After the development and district-wide implementation of HSD2's writing performance assessment in AY 1989-90, HSD2 began work on revising the K-12 reading and language arts (literacy) curriculum — the first curricular area mandated by the state to undergo the standards-based change. The cornerstone of this revision was the replacement of traditional forms of assessment with performance assessments.

The literacy PBC underwent a pilot test in a limited number of schools within the district during the winter of AY 1992-93 and was implemented district wide during AY 1993-94. The next K-12 curricular area to become performance-based will be science and math (scheduled to be piloted in a limited number of schools during AY 1995-96).

History of Reform

HSD2's adoption of performance assessments as the centerpiece of its literacy PBC grew out of teacher reaction to a declaration of HSD2's mission statement by the past superintendent. After the statement was distributed, questions arose among school-level staff about how teachers were to operationalize the mission statement. Partially as a consequence of the lack of teacher input and partially as a consequence of the questions that were raised about the district mission statement, every school in the district developed its own set of outcome indicators for students (including students with disabilities). The commonalties among the outcome indicators became obvious and a process of district-wide consensus building was begun that resulted in the adoption of the district's *13 significant student outcomes* (see Appendix A). The next step was to decide how to measure movement toward outcomes, since all involved realized that they could not be evaluated by traditional assessments.

It was not until four influential district- and school-level actors (the Director of Elementary Education, the Director of Secondary Education, the District Coordinator of Research and Evaluation, and one McGary Elementary School principal) attended a series of training sessions with Grant Wiggins of Centers on Learning, Assessment, and School Structure and Rick Stiggins of The Assessment Institute that a solution to HSD2's measurement problem was found. The four key actors, armed with the belief that performance assessments were the way in which teachers district wide could teach to and assess movement toward the 13 significant student outcomes, proposed and implemented *a four-pronged approach* to support the adoption of performance assessments in HSD2:

- The establishment of *Groups of Worriers* (comprised of district- and school-level staff) to report on concerns raised about the significant student outcomes[2];

- The establishment of the *Assessment Academy*, to heighten awareness about the purpose and the use of performance assessments among educators and administrators in HSD2;

- The development and implementation of K-12 performance-based curricula (PBC), the cornerstone of which is performance assessment; and

- The development and implementation of portfolio assessments and exit interviews as a requirement for high school graduation.

Development of the Literacy PBC and Assessments

As stated earlier, the development and implementation of the literacy PBC and performance assessments represents a break from past practice in HSD2. The new curriculum focuses on student outcomes and on performance assessments, and so meets the requirements of the state's Task Force on Student Standards and Assessment. Every teacher who taught reading and language arts in HSD2 was involved in the process of reforming the literacy curriculum. The establishment of that curriculum was guided by the following six questions[3]:

- What should our students know and be able to do?

- How will we measure what they have learned?

- To what degree or level did they learn?

- What is the best way to teach students?

- What is the best way to organize teachers, students, and content to make certain that outcomes are achieved?

- What books and equipment will best help students to learn?

During the curriculum review process, and after attending training on performance assessments, members of the reading and language arts district curriculum committee developed performance assessments and benchmarks for student achievement. The literacy PBC was piloted during AY 1992-93, underwent major revisions in the tasks and scoring rubrics during the following summer, and was implemented district wide in AY 1993-94.

---

[2] The *Worry Group Reports* were published on August 17, 1992.

[3] Excerpted from *Assessment Academy* materials.

Purpose of PBC

The purpose of PBC[4] in literacy is to:

- Set outcome achievement targets for students, classes, and grade levels;
- Conduct ongoing assessment throughout the year (which mirrors end-of-year assessments);
- Design goal-setting and improvement plans for students requiring special support;
- Design and carry out lesson plans that continuously teach outcomes and target indicators;
- Teach students to become self-assessors by providing them with ongoing, specific feedback, using rubrics, not just grades and percentages; and
- Insure that each unit of study has clearly stated outcomes and a means to assess results.

Characteristics of the Literacy PBC and Assessments

District-wide implementation of the K - 12 literacy PBC began with its distribution in September of 1993. The implementation plan is ". . . designed as a living document which is open to revision and change as more clear and authentic tasks and scoring rubrics for assessment, as well as additional components, are suggested and developed."[5] The plan contains:

- Reading/Language Arts Beliefs Statements;
- Reading/Language Arts Direction Statement;
- Quality Indicators (teacher behaviors and elements to look for in the classroom, if the curriculum is being implemented well);
- Description of components of the curriculum; and,
- Performance assessment tasks by grade level.

Each Language Arts teacher is to teach three to five tasks of the five that are included in the literacy curriculum for their grade level, and principals are encouraged to ask to see the results on these tasks. Tasks are to be administered throughout the year, so that teachers are able to teach, receive feedback, and reflect upon the assessment process. One principal's comment on this process was, "I don't see how you could do this any other way."

---

[4] Paraphrased from materials distributed during performance assessment training and included in HSD2's literacy PBC.

[5] Harrison School District 2, *Literacy Curriculum: General Information*, September 1993.

One of the five performance tasks (predetermined by the district) is utilized for school-building level accountability purposes. The classroom teacher is in charge of administering, scoring, and reporting students scores (i.e., number of students at the four proficiency levels) to the district. (Some of the tasks also are scored by teacher panels.) The assessment is not "high-stakes." Instead, it is to provide instructional direction. (As one principal noted, "Assessment really does drive instruction.") The district then reports back to the schools, providing aggregate data by grade-level, indicating students' strengths and weaknesses by grade-levels. (Appendix B is a sample assessment task, including scoring rubrics designated to be used for accountability and student self-assessment purposes.)

PBC Assessment Tasks

Literacy assessment tasks for each grade level in the areas of reading, writing, and speaking are part of the PBC curriculum. Each performance task is keyed to a target indicator and also lists the enabling processes and developmental skills required to perform the task. Target indicators provide evidence that students are moving toward achievement of the *13 significant student outcomes* appropriate for the developmental level of students at that grade level. For example, one writing performance task for third grade students is keyed to using the writing process to communicate thought, feelings, and ideas, and the significant outcome it is keyed to is *Literacy: Writing*.

The performance task is the activity that students complete to demonstrate they understand the ideas and can demonstrate the skills identified in the target indicator. In many cases, especially at the elementary level, the literacy assessment tasks have a substantive focus that overlaps with other curricular areas, such as science and history, thereby encouraging cross-disciplinary instruction. Tasks vary in their complexity and in the time required for completion (some last for several weeks). For example, one two-day task requires third grade students to write a fictional newspaper story on an animal at the local zoo. On the first day, students complete a rough draft within 50 minutes, and then work with other students as "editing partner" for two 10 minute sessions. On the second day, students again work for two 10 minute sessions with "editing partners" to finalize their papers. By working on this task, students demonstrate their knowledge of the life cycle of an animal, including a description of the first year of the animal's life, what the animal eats, how it finds its food, and where it lives. Additionally, through this task, each student demonstrates his or her skill at writing a "human interest" story.

The *enabling processes/developmental skills* are defined as the knowledge, skills, and attitudes that students must possess in order to complete the performance task and to demonstrate achievement of the target indicator. The skills represent the focus and structure of the instructional strategies and curriculum at the grade level (e.g., third graders should "use who, what, when, where, and why facts" in their newspaper stories).

Scoring Rubrics

Two scoring rubrics for each assessment task are included in the literacy PBC: the teacher scoring rubric and the student scoring rubric. The essence of the two rubrics is the same, but the student rubric is worded in the first person to reflect the fact that students are performing the task and assessing it. The student rubrics are provided to each student by a teacher at the outset of every task.

Rubrics are in a tabular format. The rows represent the task's specific criteria and the columns represent the four levels of competency adopted by the district ("In Progress," "Basic," "Proficient," and "Advanced"). Each cell of the table contains a definition of what a student's work would look like for one specific criterion at one specific level of competency. In the case of an assessment task with a written product (such as a fictional newspaper article), one criterion of that products success may be *the level of engagement of the reader*. An "In Progress" assessment of a student's newspaper article, based on that criterion, might be: "The story is difficult for the reader to understand." An "Advanced" assessment might be: "The newspaper story provides a high level of reader interest and engagement." (Sample scoring rubrics can be found in Appendix B.) Because numerous criteria are simultaneously judged, a individual student can meet the basic standard on a scoring rubric for some dimensions and the proficient standard for other dimensions.

The Process of Assessment

A Language Arts teacher employing the HSD2 literacy PBC chooses one of the five performance tasks designated for students at the appropriate level and prepares his or her students for that task. The teacher first explains the goal of the task (e.g., to prepare an illustrated book review for next year's class so that the next class may select favorite books of this year's class). The teacher then goes over the student scoring rubric with his or her students, making clear the different criteria to be judged and the different levels of achievement that can be attained. While students perform the task, the teacher frequently refers back to the student scoring rubric and asks his or her students to reflect on their performance. The teacher sometimes also asks students to perform peer evaluations of one another's work (based upon the student scoring rubric). At the completion of the task, the teacher assigns each member of the class a score on each dimension of the teacher scoring rubric (based upon each student's final product and, perhaps, an added final consultation with the student during a brief student-teacher conference).

Grades and Performance Assessments

PBC tasks are scored by teachers, by panels consisting of administrators, parents, teachers, and community members, and by students working in groups (depending upon the nature of the task and the grade level of the students). By the end of the grading period, the teacher assigns the student a traditional grade, often based on their performance on the PBC tasks.

One district-level informant said that performance assessments as an element of PBC are not the "entire answer" to educational excellence, and that they will not replace the traditional grading system. While scoring rubrics may offer more information about the performance of an individual student against a standard, the traditional grading system allows teachers to account for effort and other intangibles that are not measured by scoring rubrics. The teacher also emphasized that performance assessments are *not reliable, nor valid enough to make high stakes decisions. I'm not sure I would ever retain a kid based upon these types of assessments,. . . especially since feasibility issues like how teachers get help and how the tasks are scored are not standardized across the district.*

Students with Disabilities

Teachers provide extra support for students with disabilities, but that does not alter HSD2 standards (including time constraints on performance assessments). Given the importance of this issue, HSD2 has put a priority on the training of building-based special education coaches — a move which only serves to reinforce the district's commitment to the literacy PBC. As one district-level informant said, "We have the most to learn about the feasibility's and logistics from our most difficult students. For example, we are aligning Individualized Education Plans with the district's 13 significant student outcomes."

Evaluation of the PBC

Although there has still been no formal evaluation of the literacy PBC, an evaluation is expected in the near future. At that point, the reliability and validity issues will be systematically addressed.[6]

Staff Development and Financial Support

Staff Development

During AY 1992-93, HSD2, through its *Assessment Academy*, trained all district teachers in the use of the literacy PBC. Given that very few people within the district have expertise in performance assessments, district-level officials acknowledge that this training is key to successful implementation.[7] The training for the literacy PBC was conducted in three sessions for each grade level, one day of which was contracted in-service, and was conducted by the four individuals who spearheaded the assessment reform in the district (i.e., the Director of Elementary Education, the Director of Secondary Education, the District Coordinator of Research and Evaluation, and the McGary Elementary School principal). The purpose of the training was to help teachers ". . . gain knowledge and expertise in the development, administration, and interpretation of performance-based student assessments.[1] Overall, teachers feel that the district has done a ". . . very good job of allowing teachers access to adequate support, such as training and materials; and as one teacher said, "I haven't had to go outside the district to get the support I need." The description in Exhibit II of a professional development session for Colorado College students is illustrative of the type of professional development sessions that were provided to HSD2 teachers. This session was devoted to building awareness of performance based assessment systems.

Although the idea of performance assessment was not entirely new to the Colorado College students, some of the vocabulary and how to fit a systematic performance assessment program into the classroom were new ideas.

As of September 1995, no formal evaluations had been conducted. Although some dates were collected for determining interater reliability, no results were available.

Some HSD2 personnel were trained by Grant Wiggins and other performance assessment experts.

Harrison Academy 1992 Brochure.

EXHIBIT II

Assessment Academy

Forty Colorado College students came to HSD2 to take advantage of the type of professional development sessions the *Assessment Academy* provides to HSD2 teachers. McGary Elementary School Principal and the Director of Elementary Education led the session. The students were particularly interested in learning how to structure the teaching of science at the elementary school level and how to assess student learning.

McGary Principal covered the three questions she posed at the beginning of the session:

- What is performance Assessment?
- Why would I do it?
- What does it look like?

Throughout the presentation, McGary Principal stressed that performance assessments can be thought of as "value added" assessments. That is, content does not have to be compromised in the shift from multiple-choice tests to performance assessments. She then showed a video of a science assessment. (One student's response to the video was, this is no different from what we do.)

Colorado College students brought up several concerns. These were:

- How can I make assessment rubrics?
- How can I communicate with parents about the rubrics?
- How can I translate rubric scores into letter grades?
- It will be time consuming. How can I cover content?
- What can I do with the information from assessments?
- How much of the old testing system should I keep?

In addition to the traditional method of providing staff development, HSD2 adopted a peer coaching model to help staff as they implemented the literacy PBC during AY 1993-94. Chapter 2 funds were used to support one coach per secondary building and two coaches per elementary building with training, 5 release days, and a stipend of $300. The literacy coaches, who were teachers selected by HSD2 principals from a pool of applicants, receive ongoing training once a month on various topics. The training is not in performance assessments per se, but on related skills, such as facilitation, that enable peer coaches to encourage and support the adoption of performance assessments. Their specific responsibilities (HSD2 Literacy Curriculum Implementation Plan 1993-94) include:

- Serve as building resource to peers as they implement Literacy PBC;
- Serve as a communication link between building staff and central office curriculum persons;

- Conduct peer observations and conferences (by request) related to implementation of Literacy PBC;

- Assist grade level teachers in scheduling assessments and compiling assessment data;

- Assist in designing, modeling, and delivering Literacy PBC curriculum information to parents and community; and

- Participation in summer curriculum review work focusing on curriculum alignment and task revisions.

Financial Resources

HSD2 has a long tradition of staff development and has received national recognition for its efforts, which were at least in part driven by spending $40,000 a year during AY 1992-93 and 1993-94 on the development of performance assessments. During that same period, $500 a year was spent on assessment materials, and $6000 per year was spent scoring performance assessments.

One district-level informant explained that:

> *. . . teachers wouldn't probably say that money is an issue. [The Directors of Elementary and Secondary Education] . . . are in charge of the instructional money for the district, and they recognize that money gets you what you need. They have always been generous and far-sighted in budgeting for curriculum and assessment development — trading funds with other accounts if necessary.*

Other District-Level Tests, Assessments, and Reforms

The Iowa Test of Basic Skills (ITBS) is administered to all HSD2 students in grades 3, 5, 6, 8, and 10. Since AY 1988-89, a performance-based writing assessment has been administered to students each spring at grades 1 through 5, and at grades 8 and 11; the assessment is scored by teachers (including substitute and retired teachers). Students at each level are released one day to allow teachers district-wide to score at their grade levels.

This HSD2 assessment was adapted from the state-level writing assessment. During the first year of district-wide implementation of the writing assessment, a trainer from the Colorado State Department of Education was hired to help implement the assessment. Since that time, HSD2 has re-trained its own personnel to score the assessment.

In addition to the above process, two major reform efforts are being undertaken at the district level: (1) the inclusion of special education students in general education classrooms (AY 1993-94 is the first year of full, district-wide inclusion); and (2) school-based management (which also is in its infancy — AY 1992-93 is its first year of implementation). Neither of these reforms is perceived to be related to the HSD2's move to adopt PBC.

School Context

> *Our mission at McGary School is to model and nurture growth of lifelong learners who value each persons uniqueness and who face the world's challenges with confidence and resourcefulness.*
>
> Mission Statement of McGary Elementary School
> posted in fourth grade classroom

A parent interviewed for this study observed. McGary Elementary School is ". . . not your typical elementary school. It is way above average. The general feeling expressed by district- and school-level informants was that McGary is an innovative school with a dynamic principal — a leader who also happens to have been one of the key figures in the adoption of performance assessments in the district and who has also conducted performance assessment training. The principal, who has worked in HSD2 for 15 years and has maintained a part-time appointment for the past three years at the district-level, is taking advantage of early retirement after AY 1993-94. While she asserts that the reforms are a part of them, not me, others expressed doubts that all of the seeds of her labor will bear fruit once she has left the school.

The principal herself admits that she encouraged McGary to become the school in HSD2 that has most proactively embraced the new literacy curriculum. The school also is the furthest ahead in articulating and defining scoring rubrics and student proficiency levels beyond the literacy curriculum. For example, McGary's teachers have developed scoring rubrics in several other areas: (1) a learning rubric on such measures as *ready to learn, effort and participation, respectful speech and attitude, coping with frustration, quality of my work*, and *handling my problems*); (2) a peer mediator rubric for those in the peer mediation program sponsored by the school social worker (which scores students on such measures as *understanding the concept of conflict, listening, perception/ conflict analysis, facilitation process*, and *problem-solving skills*); (3) a sportsmanship rubric for physical education class (which scores students on such measures as *understanding, positive attitude, winning and losing, respect for others*, and *respect for myself*); and (4) and an artist rubric for students in art class (which scores students on such measures as *center of interest, balance, craftsmanship, creativity*, and *historical/cultural significance*). (Appendix C contains the peer mediation rubric.) Furthermore, McGary staff are in the process of developing a parenting rubric.

For most other schools in HSD2 inclusion was a real challenge during AY 1993-94, but it was almost the reverse at McGary. McGary fully included all children with disabilities in their classrooms three years prior to the 1993-94 school year, at a time when they housed the district program for all elementary students who were emotionally disturbed. Now that those students have returned to their local schools, McGary has taken on serving a fewer number of students with disabilities.

Relation to Other Reforms and Programs

For AY 1993-94, each School Leadership Council (composed of teacher representatives from each grade level, a special education representative, representatives of other instructional specialists, the principal, and representatives of parents) was faced with determining its school's role in the implementation of PBC. During that year, McGary's School Leadership Council was still defining its overall role and its processes. The Council's first task was to work toward encouraging the increased involvement of parents in the education of their children.

Each semester, after the first two weeks of school, a goal-setting conference is held for each student and is attended by the student, parent, teacher, and if appropriate counselor/social worker. The purpose of this student-led meeting is to set goals that enable students to become better learners, readers, writers, and mathematicians.

After students receive their first report card, only struggling or new students have a goal-setting conference at the start of the second quarter. All students attend yet another goal-setting conference at the beginning of the second semester (and, again, only new or struggling students have a conference at the beginning of the fourth quarter). McGary's principal said that the conferences are deemed "very powerful and are the most highly regarded reform that the school has undertaken — at least as far as parents are concerned. The value of the conferences to parents, according to one parent, is that they let ". . . parents . . . know exactly what their kids know".

Although many student accomplishments are acknowledged in the classroom, the McGary Student Excellence Program was established to celebrate school-wide student achievement and accomplishment. The program presents students with awards in seven areas of achievement for grades one through five, which are intended to encourage the achievement of the highest standards expressed in the many rubrics McGary students use every day. These rewards include *At-home Reading Award* and *Reader in the Classroom Award.*

The principal of McGary also established joint planning periods for staff so that all of the teachers in the school at the same grade level have their planning periods (three to five per week) at the same time, which facilitates the teaching of one theme across subjects and allows teachers time to discuss and prepare together for performance tasks. Students also are released early on Fridays to give teachers extra planning time.

Impact of PBC and Performance Assessments

The impact of and reactions to PBC and performance assessments at McGary has been largely positive. These are discussed below.

Impact on Teachers

One teacher and literacy coach related that ". . . there was initially a mixed reaction among the faculty about the move to PBC and performance assessments. There were clearly some skeptics and some early adopters, but there was respect for the process of implementation undertaken by the district and for our principal.

From the beginning of the process in the fall of 1990, teachers knew that "nothing would be held sacred" as HSD2 began to reform the literacy curriculum. Therefore, many teachers were "afraid" of the new process. Indeed, as one teacher said, "Teachers are comfortable with traditional teaching," and some teachers feel that this change is ". . . hard, and just one more thing . . . [they] have to do."

The movement toward performance assessments included classroom management issues ("How do I handle 52 students"?), grading issues ("It seems like double the work."), data issues ("I don't know what to do with all the information I already get."), and instructional issues ("It is hard to meet individualized needs."). Indeed, barriers to the adoption of performance assessments, all explicitly mentioned during the HSD2 Assessment Academy, include: time, class size, parental expectations for familiarity, teachers who are resistant to change, lack of administrative support, lack of role models, administrative expectations for familiarity, pressure on teachers to conform to the "chalkboard" lecture, and pressure from the media. For all of these reasons, one of the main goals of the assessment academy was to supply teachers with a clear rationale for adopting performance assessments. In short, teachers had to be assured that *they assess only what is valuable*.

When performance assessments were first implemented district-wide (in the fall of 1993), teachers experienced some confusion regarding the semantics of the performance tasks and the fine tuning of those tasks, and some teachers felt they were trying to assess too many elements at the same time. One teacher complained that as part of a performance task, individual student listening skills had to be assessed while other students made their oral presentations.

While most teachers realize that the shift to PBC and performance assessments is not perfect, they are not refusing to become engaged with the process. In fact, one teacher expressed a desire to expand the implementation of performance assessments beyond the literacy PBC.

Impact on Curriculum and Instruction

Most teachers agreed that PBC and performance assessments changed their pedagogical techniques for the better; and that the assessments changed the very nature of classroom instruction. Teachers now believe that performance assessments give them more information about their students, give them a better idea of what and why they are teaching, and give students an explicit goal and the means of achieving it. More than one teacher shared the vision that ". . . large changes to instruction will result from this movement;" and one teacher emphatically expressed that "performance assessments have helped me to become a better teacher."

Teachers at McGary attempted to do something they had never done before (i.e., stage a public academic performance for parents), because, as one teacher said,

> *When you perform, you practice. The quality of work improves. We all respect a good performance, but we hardly ever see an academic performance. We wanted our community to know that we respect academic performance.*

As part of a unit on Westward expansion and in preparation for the performance, each student in the fifth grade wrote a research article or diary entry on the topic of the Oregon Trail. The night of the performance, the teachers set up mock campfires with groups of five to six students (dressed appropriately in pioneer attire) telling stories based upon their research article or diary entry. As one teacher said, ". . . we weren't sure exactly what would happen, but we got huge support from all of the other teachers. We even gave up teaching *Call of the Wild* so that we could have all the time we needed to pull it off."

Parents were invited to come to the school after work, not only to watch, but to participate in the scoring of their child's performance (based upon the rubric for the task). To allow students and parents to review the student's performance, each of the performances was videotaped. The response was "tremendous." As one teacher said, "It was worth it . . . to see the excitement the kids felt. I am sold on it." As of the spring of 1994, the fifth grade staff at McGary had gone so far as to plan inservices for other schools on public academic performances.

Impact on Students

Teachers suggest that PBC and assessment rubrics have helped to clarify standards for their students. Elementary students themselves expressed that PBC and performance assessments can "help you get a better grade than you used to" and "can tell you how to complete your work." One student said that she liked the scoring rubrics, as they made her "stories better," she said, so that she can now write an "advanced-level story." She said, "I look at the rubric to make sure everything is there. Mom helps me to identify where I am on the rubric sometimes."

Teachers, however, are even more enthusiastic than students about the impact on students of PBC and performance assessments. One teacher was pleasantly shocked to see how *student-empowering* the scoring rubrics were. "I have never before," she said, "had a student come to me with a B and ask, 'How do I get an A'?"

One teacher, however, did qualify the enthusiasms of her peers by stating that, "It is too early too tell if . . . [students] have improved performance." Additionally, some teachers expressed concern that students with disabilities included in their local classroom may not all be able to handle the performance tasks. "Still, " she said, "for our school, it is better to have the performance assessments than not to have them."

Impact on Parents and Community Members

One parent said, "The performance assessments have been well accepted by 98 percent of the parents, but I wouldn't want to see the traditional grading system go by the wayside." Parents want to know whether their child is "an 'A' or a 'B' or what?" In general, parents and community members are comfortable with the change and feel that HSD2 is "adding something, not taking something away" with the implementation of performance assessments. The same parent informer said that performance assessments ". . . have definitely increased the involvement of most parents." This is important, as transient parents may tend to be less involved with school activities than families with a long-term "investment in the community."

The school board member was of the opinion that PBC is an "exciting story about how teachers think about their students." He also said that the entire board is not excited about the performance assessments, even though it is not fully aware of what is happening at HSD2. However, the board likes the "go slow" approach and deals with skepticism through discussions. He said that the community needs to be brought along, because it often opposes reforms as a result of feeling left out (and not necessarily because it feels that the reforms are not academically sound). One of the most pressing issues the school communtiy has to deal with, he said, is the disparity in achievement between minority and white students.

Future Plans

As one teacher said: "The [curriculum reform] process is going to take a long time, perhaps 10 years before it reaches some semi-final form." So far, teachers and parents see the assessment effort as a "value-added" experience that has great potential for realizing the 13 significant student outcomes, but not until the distant future. Indeed, the lessons of Littleton, Colorado are well-heeded in HSD2.

Notwithstanding, there is much that district- and school-level informants would still like to "see happen." They would like to see more thought go into the inclusion of students with disabilities; they would like to see a better articulation between the curriculum and assessment of different grades; and they would like to see the continued refinement and adoption of performance assessments. Some expressed they would like to continue expanding the application of performance assessments across all areas of learning, in and out of school, and over summer breaks.

HSD2 plans to follow the state-mandated plan to move to standards-based education. They will have piloted the science and math curriculum by AY 1995-96, and the history and geography curriculum will be piloted by AY 1997-98. After each school year, during the summer, curricula and assessment tasks are modified and improved, and teachers receive additional training and time to reflect upon their instructional practices.

Epilogue

During the summer of 1994, the HSD2 staff met to discuss and address several issues with the PBC and performance assessments. They simplified the language of the student scoring rubrics, spent time articulating the curriculum across grade levels (although that work is was not finished), and drafted guidelines regarding how much support to provide to students on performance tasks. The new literacy PBC was reprinted and distributed to all HSD2 teachers.

Conclusions

With the adoption of performance-based assessments as the cornerstone of their new PBC, HSD2 has changed the very way they view classroom instruction. No longer do multiple choice tests determine what is taught in the Language Arts classes. What happens in the classroom is now determined by performance assessments — performance assessments developed within the district, *by* teachers *for* teachers.

HSD2 district-level staff established the rules for the current reform; they set the standards; they set the method of assessment; and they provided the support, both emotional and financial, for change. It is teachers, though, at each school and in each classroom, who are ultimately responsible for the success or failure of this reform. While HSD2 has been ahead of the state with its mandate for standards-based education, it has allowed individual teachers the freedom to determine what that mandate actually means.

No doubt the retirement of the principal of McGary Elementary School will reverberate throughout every school in the district. A strong proponent of reform, the principal was a key actor in bringing change she supported to HSD2. Her leaving will no doubt make reform efforts that much more demanding.

One second grade teacher summed up the feelings among HSD2 personnel about those ongoing reform efforts: "I think it will be worth it. I'm excited to be focusing on what we will be doing next year. These are stressful changes, but we are getting support." HSD2, in short, is prepared for positive reform.

APPENDIX A

13 Significant Student Outcomes

HARRISON SCHOOL DISTRICT

Significant Student Outcomes

LITERACY

Reading:

Students will be able to construct meaning when reading in all subject areas for the purposes of becoming informed, performing a task, and enjoying literature.

Writing:

Students will communicate in writing to multiple audiences for the purposes of informing, persuading, organizing, and providing enjoyment.

Listening & Speaking:

Students will be able to listen with understanding and speak with clarity and purpose.

PROBLEM SOLVING

Students will use problem solving strategies both independently and collaboratively in academic areas and interpersonal relationships.

MATHEMATIC AND SCIENTIFIC REASONING

Students will apply mathematic and scientific principles and processes for the purpose of solving problems.

INTERDEPENDENCE/DIVERSITY

Students will demonstrate an understanding of the concepts of personal and environmental interdependence, and an appreciation of cultural diversity in a global community.

Students will use a historical perspective to understand and apply fundamental American principles that lead to active, responsible citizenship.

EMPLOYABILITY

Students will develop skills and attitudes necessary for successful employment and will make career choices based on self-analysis and personal understanding.

WELLNESS

Students will apply knowledge and will develop healthy habits of positive self-esteem and wellness for the purposes of making responsible lifestyle choices.

TECHNOLOGICAL LITERACY

Students will use technology for life-long learning and will develop an understanding of technology's impact on the quality of life.

FINE ARTS

Students will be active participants in exploring and developing individualized skill in creative expression and will be able to respond to the creative work of others.

LIFE-LONG LEARNING

Students will actively pursue and apply self management, organizational skills, and learning as a life-long process.

PERSONAL/SOCIAL

Students will demonstrate self-discipline and apply knowledge and standards of conduct for the purpose of ethical decision-making.

APPENDIX B

Sample Literacy Task and Rubric

DIRECTIONS FOR ADMINISTERING TASK

The task involves some pre-teaching and/or review of some writing skills. In order for this to be a successful assessment task, students must understand the difference between a report (including as much information about a specific animal) and a newspaper article (including a human element in the story while informing the reader of the 5W's).

Also, standards for editing need to be reviewed if these are not already an integral part of daily instruction in the Writer's Workshop.

It might be helpful to do a mini-unit on newspapers prior to the introduction of this task. Creativity is expanded when the students are free to include a headline, byline, dateline, logo, cutline and include a picture of their subject. (See Newspaper Packet)

Students will brainstorm and select animals to be featured in their articles. Students will identify information that they want to include in their articles through research, interview, observation, etc. This pre-writing activity is essential to producing a quality article.

First Day

As a reporter you are very used to deadlines! This assignment will need to be completed in two sessions. The first session has a deadline of 50 minutes for your rough draft. For the second session you will work with editing partners for two 10-minute sessions. At the end of that time you will be asked to turn in your newspaper article to the Chief Editor - Me!

You need to keep in your head what a quality piece of writing looks like and sounds like. You may begin.

As a reporter you will also be evaluated on writing so it is very important that you edit your paper carefully after you get your main ideas down on paper. The grader will be looking at the organization of your article as well as sentence structure and spelling.

Second Day

Today your job will be to "fine tune" your piece of writing. We will be working with two student editors for approximately 10 minutes. It will be your job to help each other with all corrections you feel you wish to do to your piece of writing. At the end of that time we will change partners and have another student help you with your piece. At the end of the second session I will collect your papers for evaluation.

You need to keep in your head what a quality piece of writing looks like and sounds like! You may begin.

The following directions would be given to students before and during their Writer's Workshop:

The teacher would say:
Your article should be a very well-organized and easy-to-understand story with a good beginning, middle and ending. You must be very accurate and give lots of details to make your story "come alive". You must make the reader a picture with your choice of words. Include a human interest side to your story that would make your reader curious and informed about your topic. You must include the 5 w's (who, what, where, when and why) in your story. You must edit your paper very carefully for spelling and word order.

These directions need to be run off for each student.

POST WRITING INSTRUCTIONS

For this part of the task you, the teacher, will need a copy of the scoring rubric. The rubric contains important elements for the various levels of proficiency. If you can solicit help from an outside unbiased grader it is quite helpful. The time needed to evaluate this is longer than a normal writing assignment as you are looking for specific criteria included in the rubric. It is very helpful to review the rubric and become very familiar with it before you attempt to score. It is also helpful to do this at one sitting if possible to have more reliability.

LITERACY URRICULUM

Grade: 3

SIGNIFICANT STUDENT OUTCOME: Literacy: Writing Students will communicate in writing to multiple audiences for the purposes of informing, persuading, organizing, and providing enjoyment.

***Target Indicator # 2** Students will use the writing process to communicate feelings, thoughts and ideas. (In various types of writing.)

***Assessment Task:** **(The Cheyenne Mountain Zoo Task)** Each month the Cheyenne Mountain Zoo publishes a newspaper to inform children and their families about special events taking place at the zoo. The newspaper features an article called "Who's New at the Zoo?" The article reports on new animals brought to the zoo or new animals' births, including the who, what, when, where and why facts.

As a junior reporter for the zoo news, your task is to select an animal and introduce the animal to your readers. Your article should contain an accurate description of the life cycle of your selected animal, along with a description of the first year of life for the newborn of this animal. Your readers will want to know what this animal eats, how it finds food and where it lives. Your article should also contain the human interest side of the story - a new birth, an animal on loan to another zoo for display or research, or perhaps the death of an animal that communicates feelings of the zoo keeper.

ENABLING PROCESSES DEVELOPMENTAL SKILLS
Letters:
- Identify and apply 5 parts of friendly letter
- Write invitations and thank-you notes
Reports:
- Write to give information: book reports, reviews, scientific observation, recipes, how-to reports, beginning level research projects and other descriptive writing
Journals:
- Keep personal record of experiences and feelings
News Articles:
- Identify and begin to use who, what, when, where, why facts
Poetry:
- Create poetic image using rhyme, rhythm, tone and mood

Other possible Assessments (apply to all Enabling Processes/ Developmental Skills)
Observation
Anecdotal records
Writing samples
Conferencing
Student self-evaluation
Peer evaluation
Other:

SCORING RUBRIC ·d GRADE LITERACY

OUTCOME: Literacy: Writing **Students will communicate in writing to multiple audiences for the purposes of informing, persuading, organizing and providing enjoyment.**

TASK: (See "Cheyenne Mountain Zoo" Task)

	STANDARDS: The levels at which students perform the task.			
	In Progress	**Basic**	**Proficient**	**Advanced**
CRITERIA: The specific behaviors, products and qualities we look for in judging student work.	There is little or no evidence of organization. Sentences are isolated and do not create a story. Limited factual information is given. Only 1 or 2 of the 5 W's was present. The story is difficult for the reader to understand. Sentence structure, grammar, mechanics and spelling are consistently weak. No human interest element is evident.	The story is loosely organized, and the overall presentation is choppy and somewhat confusing without a logical flow from beginning to end. Factual information is given, but with some inaccuracies. 3 of the 5 W's were included. The story engages the reader in a limited manner. Sentence structure, grammar, mechanics and spelling are noted to have weaknesses and errors. Minimal human interest is evident.	Narration may have minor organizational flaws (digressions or repetitions), but the ideas flow logically throughout most of the story. Supporting information is basically accurate, with some details and elaboration. 4 of the 5 W's were included. The reader is engaged in the newspaper story. Sentence structure, grammar, mechanics and spelling are generally correct. The article included a human interest element.	Narration is clearly organized and focused. Ideas flow logically from beginning to end. Information given is factual and accurate and highly detailed. The reporter included who, what, where, when,or why using vivid language. The newspaper story provides a high level of interest and engagement by the reader. Sentence structure is varied, and grammar, mechanics and spelling are consistently correct. A definite, clear and engaging human interest element is included.

3rd grade rubric / 7

BEST COPY AVAILABLE

STUDENT SCORING RUBRIC - 3rd GRADE LITERACY

OUTCOME: **LITERACY: Writing** **Students will communicate in writing to multiple audiences for the purposes of informing, persuading, organizing, and providing enjoyment.**

TASK: **(See Cheyenne Mountain Zoo Task)**

	STANDARDS: The levels at which students perform the task.			
	In Progress	**Basic**	**Proficient**	**Advanced**
CRITERIA: The specific behaviors, products and qualities we look for in judging student work.	• My article does not really make a story. I don't have much factual information. • I only included 1 or 2 of the 5 W's. • It is difficult to understand. • Spelling, punctuation and grammar are not too good. • I did not include human interest element.	• My article is choppy and does not flow like a story. I included some information. I am not sure if all the information is true. • I included 3 of the 5 W's. • The reader understood most of my article. • The spelling, sentences and punctuation are weak. • I included only a little human interest element.	• My article flows and is easy to follow. I have included accurate information and some elaborations. • I included 4 of the 5 W's. • The reader enjoyed my article. • The sentences are varied. Grammar, spelling and punctuation are basically correct. • I included a human interest element.	• My article is clearly organized and is easy to follow. The information is factual, accurate and highly detailed. • I included all of the 5 W's. • The reader was very interested in my article. • The sentences are varied and mechanics and spelling are consistently correct. • The human interest element is clear and engaging.

APPENDIX C

Peer Mediation Rubric

SCORING RUBRIC FOR STUDENT MEDIATOR PROGRAM

Student Mediator Program	STANDARDS: The levels at which students perform the task.		
	Beginning	**Intermediate**	**Advanced**
Understanding the Concept of Conflict	When I am discussing conflict, I am able to list various types of conflict.	As a mediator, I acknowledge that conflict can be good and bad.	As a mediator, I know that conflict is neither good nor bad; it simply exists. Conflict creates change which is a growing process.
Listening (Communication) Skills	As a mediator, I try to listen to both parties.	As a mediator, I often use effective listening skills to foster communication.	As a mediator, I consistently use the five active listening skills of : 1. attending; 2. nonverbal communication; 3. para-phrasing; 4. summarizing; and 5. "I" messages.
Perception/ Conflict Analysis	As a mediator, I try not to take sides when mediating.	As a mediator, I understand both sides of the problem.	As a mediator, I am able to understand both sides of the problem without bias by acknowledging the disputants' perspectives.
Facilitation Process	As a mediator, I am able to orchestrate the mediation process by following the four basic rules: 1. to tell the truth; 2. No name calling; 3. try to solve the problem and 4. no interrupting.	As a mediator, I use de-escalators, as well as the four basic mediation rules. I have these rules memorized.	As a mediator, I facilitate the mediation process by using the following skills: 1. listening both verbal & nonverbal; 2. moving from the scene of conflict; 3. using de-escalators; 4. turn taking; 5. control/structure process; and 6. follow basic mediation process.
Problem-Solving Skills	As a mediator, I strive to reach good resolutions.	As a mediator, I assist disputants in reaching compromises by focusing on disputants needs.	As a mediator, I assist disputants by reaching win/win solutions by addressing and drawing disputants from positions to interests.

South Brunswick's Sixth Grade Research Performance Assessment: Windermere Elementary School

June 2-3, 1994

SOUTH BRUNSWICK'S SIXTH GRADE RESEARCH PERFORMANCE ASSESSMENT: WINDERMERE ELEMENTARY SCHOOL

Introduction

The South Brunswick Township Public Schools developed the Sixth Grade Research Performance Assessment after coming to the realization that the district's students were leaving elementary school without the research skills the district wanted them to have. First piloted in 1991, the assessment has evolved each year since, becoming a widely popular feature of the district's educational program. This study focuses on the development of the performance assessment and its implementation at Windermere Elementary School.

Windermere Elementary School is one of seven elementary schools in the district. During the 1993-94 school year, the school served 500 students in grades kindergarten through six. Sixty-nine percent of those students were white, 12 percent were African American, 16 percent were Asian-American, and 3 percent were Hispanic. Nine percent of students qualified for the U.S. Department of Agriculture's free or reduced-price lunch program. The school staff includes 27 teachers, one administrator, one librarian, one part-time guidance counselor, and one secretary.

Teachers at Windermere and school and district administrators alike say that the Sixth Grade Research Performance Assessment has resulted in changes in instruction in the upper elementary grades and in increased collaboration between teachers and school librarians. The district embraces the concept of "resource-based instruction," an idea which bolsters the Sixth Grade Research Performance Assessment and the instructional emphases it encourages. Thus the assessment, coupled with resource-based teaching, is intended both to measure students' progress toward attaining research skills and to provide a structure for teachers to teach those skills.

Participants

In addition to the South Brunswick Township Public Schools' Assistant Superintendent for Curriculum and Instruction, those individuals associated with the Windermere Elementary School who participated in this study are listed in Exhibit I.

EXHIBIT I

Study Participants

- Assistant Superintendent for Curriculum and Instruction
- Four teachers, three sixth-grade teachers and one resource room teacher
- School librarian
- One parent
- Four sixth-grade students

Observations

The two researchers conducting this case study participated as outside assessors for the Sixth Grade Research Performance Assessment, undergoing the training provided assessors by the district and each assessing six students' research efforts.

District Context

The South Brunswick Township Public Schools (South Brunswick) operate seven elementary schools, one middle school, and one high school in central suburban New Jersey. The district primarily serves children from middle class homes and is ethnically diverse.

Development of the Sixth Grade Research Performance Assessment

The development of the Sixth Grade Research Performance Assessment followed several events. The construction and opening of a new, state-of-the-art elementary school led South Brunswick educators to revisit the district's elementary curriculum. As a result, in about 1990 South Brunswick teachers developed a set of "outcome statements" for elementary schools — a set of expectations of what students leaving South Brunswick's elementary schools should know and be able to do. (See Appendix A for a copy of the sixth grade outcome statements.)

Upon reviewing both the outcome statements and the district's curriculum, South Brunswick teachers and administrators discovered that, although the documents emphasized research skills, students were, in all likelihood, leaving sixth grade deficient in this area.

As a result of this realization, South Brunswick developed the Sixth Grade Research Performance Assessment. The district drew on the work and ideas of four different organizations or individuals in developing its assessment, as follows:

- The district's familiarity with performance assessment in the form of literacy portfolios already in use with younger children;
- The work of Grant Wiggins of the Center on Learning, Assessment and School Structure and the Littleton, Colorado, school district in developing performance assessments;
- The theory of "resource-based instruction," a pedagogy which emphasizes the simultaneous development of research skills and mastery of content; and
- Professional assistance from the nearby Educational Testing Service.

The Sixth Grade Research Performance Assessment was piloted in 1991 in four sixth grade classrooms. In 1992 it was implemented districtwide for the first time.

Characteristics of the Sixth Grade Research Performance Assessment

The Sixth Grade Research Performance Assessment, administered in May and June each year, comprises:

- **An extended performance task**. Students have eight hours to:
  - Research a question related to the "American experience" using different sources of information (e.g., encyclopedia, books, people, CD ROM);
  - Prepare a bibliography;
  - Prepare a written report; and
  - Prepare a three to five minute oral presentation with at least one relevant visual.
- **A performance event**. On the day following the students' completion of the performance task, they must present their oral reports before five or six classmates, several fourth- and fifth-grade students, and two outside assessors.

About two weeks prior to the scheduled assessment, each sixth-grade student develops two research questions related to the "American experience." Some recent research questions have included, "What is Hillary Clinton's role as First Lady?" and "Why is George Washington called the 'Father of his Country?'" The student's teacher or the school librarian assists the student in formulating the questions, ensuring that they are researchable. All questions must be questions that can be researched given the resources available at the school; no materials may be brought in from home, and students cannot conduct any of their research outside the school building.

Prior to the assessment, the students' teacher and the librarian review the research questions and select one question for each student to work on. The students learn which of their questions they will research on the morning the assessment begins. Students work in the school library, and their teachers are not present. They may ask any adult, except their own classroom teacher, for assistance in proofreading their reports, but they may not obtain any help on content.

The Sixth Grade Research Performance Assessment is administered somewhat differentially across the district's seven elementary schools. All students have the same amount of total time (eight hours) to complete the assessment; however, at some schools the eight hours is spread over two days, while at others it is spread over four days. However, aside from this variable, the basic format of the assessment task is the same at all schools and for all students.

Scoring

Students' research projects are scored on four dimensions: the written report, the visual, the oral presentation, and a process score. Each dimension of the assessment is worth five points, for a total of 20 points possible. Appendix B includes a copy of the scales used to score the assessment tasks. The four dimensions of the assessment are as follows:

- **Written report**. Students' written reports are scored on a holistic scale, with scores ranging from *1* to *5* (the highest score possible). A score of *3* is considered a good, or passing, score. All characteristics of a score-point must be present in order for a student to obtain a score. For example, to obtain a score of *3*, students' written reports must demonstrate all of the following characteristics: "*Basically competent: satisfied requirements* — The student states the question. He/she answers the question with a small amount of supporting information. There is a basic organization but it is not always clear. The structure and mechanics are generally correct with some errors. Two different types of sources are noted." Students may earn scores between two points on the written scale; for instance, if some but not all elements of the scale for a *4* are present in a written report, a student may receive a *3.5* for his report. Finally, a student who fails to include a bibliography may score no higher than a *2* on the written report.

- **Visual**. The visual the student creates to support the oral presentation is scored using a "trait" scale; each element, or trait, is either present, partially present, or not present in the student's visual. Five items appear on the scale, and for each element a student obtains a score of *1*, *½*, or *0*. The five elements on which the visual is scored include neatness; creativity; appropriateness; enhancement of the question under study; and integration into the presentation.

- **Oral presentation**. The student's oral presentation is also scored on a trait scale with five elements. Again, elements are present, partially present, or not present in the student's presentation and are assigned scores of *1*, *½*, or *0*, respectively. Elements assessors look for in the oral presentation include an appropriate introduction; a clear and audible voice; eye contact; verbal strategies such as metaphors and colorful examples; and inclusion of supporting data to answer the research question.

- **Process**. The process the student follows during the eight hours he or she spends preparing the report, presentation, and visual is critiqued by an assessor using a five-element trait scale (again, students score on each trait a *1*, *½*, or *0*). Elements of the process on which students are scored include evidence of a plan designed to achieve the project goals; appropriateness of the plan to achieve the task; demonstration that the student knows how to locate and use resources; evidence of note taking, drafting, revising, and editing in the writing process; and evidence the student can assess the progress of his or her own work.

The Scoring Process

Students are assessed on the four dimensions in two ways. First, while the student is working on the performance tasks, an adult at the school — other than the student's teacher — rates the student on the process scale. This other adult is not one of the assessors who critiques the student's performance in the remaining three areas of the assessment. The process assessor scores the student's approach to assessment tasks by interviewing the student at some point during the eight hours of the performance task.

For the 1994 administration of the assessment, the district, following the recommendations of past assessors, did not reveal the process score to the two assessors responsible for reviewing the students' other work; assessors had indicated that they felt having knowledge of students' process scores biased their scoring of students' written, oral, and visual work.

The day after the students complete their written reports, oral presentations, and visuals, assessors from outside the school come to the school to assess the students' work. These assessors include parents, community members, and educators interested in learning more about the Sixth Grade Research Performance Assessment.

Assessors work in pairs, each pair assessing the work of about six students. Students deliver their oral presentations before the two assessors (several younger students — fifth and fourth graders — also watch the presentations to give them an idea of what they will encounter in sixth grade). Following the presentations, the assessors compare their scores for each student on the oral and visual trait scales. They discuss any points of disagreement until they come to agreement on a single score for each trait. For example, if one assessor assigned a student a *1* on the visual trait "is a well integrated part of the overall presentation," whereas the other assessor assigned the student a score of ½, the two assessors must agree on assigning the student a score of either *1* or ½.

After the students have completed their presentations and the assessors have scored students' oral and visual work, the assessors turn their attention to the students' written reports. As has been described above, the scale applied to students' written work is a five-point holistic scale from *1* to *5*; whole and half scores are possible, i.e., a student can obtain a score of *3* or *1.5* or *4.5*, for example. Again, assessors must come to agreement upon a single score for each student on the written scale. Thus, if one assessor assigns a student a *3* and the other a *4*, the assessors must decide if the student merits a *3*, a *3.5*, or a *4*.

An overall score of *12* is the threshold for a passing mark. Thus, students must achieve an average score of *3* on each of the four dimensions; however, there is no minimum score required on any one of the four dimensions. The four dimensions are weighted equally in the student's score.

The assessment format, the scoring process, and the holistic and trait scales are revisited by South Brunswick teachers and administrators each year. Prior to the 1994 administration of the assessment, teachers reviewed the scales but ultimately decided to leave them unchanged from the 1993 administration; as one teacher commented with respect to the four scales, "We haven't reached consensus yet, but teachers tend to be satisfied with it."

Evaluations of the Sixth Grade Research Performance Assessment

South Brunswick, with assistance from the Educational Testing Service, collects data on its administrations of the Sixth Grade Research Performance Assessment. Data collected include information about assessors' backgrounds, scores assigned on holistic and trait scales by individual assessor and the consensus scores reached, and students' standardized test scores. According to the district's Assistant Superintendent for Curriculum and Instruction, an evaluation of interrater agreement found it to be at about 81 percent. (Although assessors are required to agree upon a single score on each trait and on the holistic scale for the written report, the district collects from assessors copies of the scores they assigned individually *and* the scores they agreed upon. Thus, the district and ETS were able to compute interrater reliability based on these data.)

The district has also compared students' standardized test scores and their Sixth Grade Research Performance Assessment scores. Although the district found a bias on standardized tests in favor of Asian-American students, they found no such bias in the Sixth Grade Research Performance Assessment. However, according to the Assistant Superintendent for Curriculum and Instruction, the district has yet to conduct any thorough analysis of the data it is collecting on the assessment.

Training and Resource Support

The district provides training sessions for assessors to explain the holistic and trait scales and the purposes of the assessment. However, the district provides in-service professional development activities with respect to teaching research skills for sixth-grade and other elementary school teachers.

In-service Professional Development Activities

The three teachers interviewed for this case study reported no formal training with respect to the Sixth Grade Research Performance Assessment. Two of the teachers had participated in the development of the assessment and, thus, were familiar with it from the start. The third teacher, a first-year teacher, said she received no training from the district on the assessment: "I was left to approach [the librarian] on my own. If I had lacked the foresight to do so, I might have hurt my kids. A timeline for covering skills would have been helpful."

Introduction to the assessment for new teachers, however, comes through the monthly, district-wide meetings of sixth-grade teachers, beginning in January of each year. Teachers meet to discuss the upcoming research assessment. According to one teacher, they spend most of the time discussing the scoring process and scales. The purpose of these meetings is twofold: to discuss any modifications that might improve the assessment and to determine the logistics of administration.

Training of Assessors

South Brunswick invites members of the community and other interested individuals to participate in the Sixth Grade Research Performance Assessment as assessors. The district often hosts educators from other districts interested in learning about the assessment.

Assessors undergo a two-hour training session at the district immediately prior to entering the schools. The training is conducted by the district's Assistant Superintendent for Curriculum and Instruction. Assessors receive folders of materials, including:

- An overview of the assessment process;
- Copies of the holistic written scale and the oral, visual, and process trait scales;

- Five "benchmark" papers to review, one at a low level of performance, one at a high level, and three in the middle range (i.e., papers meriting scores between *3* and *4* on the holistic written scale); and

- Directions for using the scales and for recording agreed upon scores.

The trainer reviews these materials with assessors, devoting most of the training time to reading and scoring the benchmark papers using the holistic written scale. She also emphasizes that assessors should be careful not to impose their own expectations of what sixth-grade students should and should not be able to do into the scoring process; rather, assessors are asked to apply the criteria identified by the school district as the standards it expects its students to meet.

In addition to the aforementioned analysis of the assessment's interrater reliability, the Educational Testing Service has also assisted the district in analyzing feedback from the assessors, with respect to both the assessment and the training process. One example of how feedback from assessors has resulted in an amendment to the assessment has been described above: upon the recommendation of assessors, they no longer see the process scores obtained by students when the assess students' written, visual, and oral work.

The use of outside assessors and the adequacy of their training remain among the most controversial facets of the assessment, according to teachers and other school staff. According to the Windermere Elementary School librarian, "The training of assessors is one of the real problems with the assessment," and "Teachers wish assessors had more experience with kids this age." A sixth grade teacher commented that the work conducted by the outside assessors is "good [for teachers] as a reality check;" however, he also suggested that assessors do not receive enough training and that they are, on occasion, insensitive to the kids.

Resources

Administration of the Sixth Grade Research Performance Assessment is funded through the regular district budget. According to the Assistant Superintendent for Curriculum and Instruction, the only real problem the district encounters with the costs of administering the assessment are the costs of feeding breakfast and lunch to 125 assessors.

Participation in the Assessment Process

The two researchers conducting the interviews for this case study were invited to participate as assessors of the Sixth Grade Research Performance Assessment. In their capacity as assessors, the researchers participated in the training activities provided by the school district and, with partner assessors from the South Brunswick community, each assessed the performance of six sixth-grade students at a single school (Windermere Elementary School). The researchers' critiques of the training session, the scoring rubrics, and the quality of students' performances follow below.

Critique of the Training Session

The two-hour training session was, on the whole, of high quality. Assessors learned about the process of the assessment, what the assessment required of students, and the time and resource constraints faced by students. The trainer reviewed the holistic written scale and the visual and oral trait scales with the assessors. These explanations were clear and of sufficient detail to communicate to assessors the purposes and applications of the scales.

The assessors then broke into small groups (of approximately four) to read and assess five "benchmark" papers. The trainer then asked the entire group of assessors to indicate what score they had assigned to each paper in turn. The benchmark papers at the top and bottom ends of the range had high agreement among all assessors; virtually everyone agreed that one paper merited a *5* and that another deserved a *2*. However, the assessors-in-training had a harder time agreeing on scores for the three other benchmark papers, which were, according to the trainer and most of the assessors, in the "*3*-to-*4* range." The trainer emphasized that it was more important for the two assessors working together to come to agreement on a single score than it was that that score be "right." (As the assessment is of a low-stakes nature, this instruction may not be problematic. However, if the assessment is ever made high-stakes, it seems likely that the low emphasis placed on assigning the "right" score will probably have to be modified and, consequently, the training emphases changed.)

Finally, the trainer, in response to disagreement among the trainees on the right score — *3*, *3.5*, or *4* — for one benchmark paper, said, "Just trust me. It's a *4*." On the one hand, this was a valuable piece of information for assessors: they had in their folders a paper for which they knew the appropriate score was a *4*. On the other hand, this assertion was inconsistent with the trainer's earlier statement that agreement between assessors was more important than assigning the "right" score.

Critique of the Holistic Written Scale

The holistic scale applied to students' written reports has some perhaps minor but significant problems. Although the directions for application of the holistic scale require that all characteristics of a score-point be present in order for a score to be assigned, descriptions for lower- to mid-range score-points indicate that certain characteristics *may* be present in student work. For example, four of the five characteristics for a score of *2* are as follows:

- Basic information *may* be lacking;
- The writing *may* lack organization and be difficult to follow;
- There *may* be many errors of sentence structure and mechanics; and
- Sources *may* be mentioned.

Common-sensically speaking, a paper meriting a score of *2* does not meet the criteria for a score of *3* but is better than a score of *1* would suggest. Therefore, the use of the word *may* in these criteria presumably is intended to suggest that some of these elements may be present and some may not be (e.g., the paper is better than a *1* and not as good as a *3*). However, the effect of this ambiguous language is to obliterate any distinction between scores of *1.5*, *2*, and *2.5* (the effect is exacerbated by the similarity of language used to describe some criteria associated with scores of *2* and *3* — e.g., "the writing may lack organization and be difficult to follow" associated with a *2* and "there is a basic organization but it is not always clear" for a *3*). In other words, because the language used to describe the criteria for a score of *2* is vague, scores just "worse" or just "better" cannot be distinguished systematically based upon the criteria. This effect, if experienced by other assessors, presumably results in a decline in interrater reliability for scores in the lower- to mid-range of the scale.

Critique of the Quality of Reports, Visuals, and Presentations Assessed During Case Study

The written reports, visuals, and oral presentations assessed by the two researchers conducting this case study were of decidedly mediocre quality. One of the two researchers, together with the assessor who was her partner, assigned a "passing" score of *3* to only one of the six written reports she assessed. Additionally, one of the six was clearly plagiarized from the sources the student included in his bibliography. Nobody at the school — neither the librarian nor any teacher — knew what the procedure was for dealing with plagiarized work on the assessment (though, on the positive side, their ignorance here suggests that plagiarism on the assessment had never previously been a problem).

As for the visuals and the oral presentations, no student did anything other than read his or her written report for the oral presentation, and the relationship between the students' visuals and their research questions was generally tangential.

It should be stressed, however, that the librarian and the teacher of the students being assessed both said that this class had had an especially hard time formulating questions. Furthermore, the librarian suggested that, for the first time ever, the students, rather than rising to the occasion, spent most of the first day (i.e., the first four hours) of the assessment socializing instead of working (in itself a negative reflection on the students' performance, presumably revealed in their process scores). The teacher and librarian's comments about this particular group of students suggest that their performance on the assessment is probably not representative of their peers in their school.

Interaction with Other Reforms

The Sixth Grade Research Performance Assessment is not South Brunswick's only sojourn into performance assessments. The district has also introduced an early childhood portfolio, still in use, and students in grades 3 through 12 maintain a "Best Works Portfolio" that is passed on with them from year to year. The district has also experimented with several other performance assessments — in the areas of mathematics, science, and reading. The district did not continue using these other assessments because teachers and administrators were dissatisfied with them (the district may reintroduce a mathematics performance assessment at the elementary school level in the near future).

However, the Sixth Grade Research Performance Assessment — the district's response to a weak link the district's teachers identified in the elementary school curriculum — really is not an integral part of a larger education reform initiative. Rather it is an assessment designed to guide curriculum and instruction and to assess student outcomes in one area. Although the district encourages the use of performance assessments, the Sixth Grade Research Performance Assessment is a piece of the district's assessment program, but not a piece integrated into a whole.

School Context

Windermere Elementary School, located in a subdivision of middle-class homes, is surrounded by extensive playing fields and lots of trees. The school is an "open space school," with classrooms separated from each other with movable walls and bulletin boards.

Impact of the Sixth Grade Research Performance Assessment

Teachers, students, and parents at Windermere Elementary School all have positive things to say about the Sixth Grade Research Performance Assessment. In addition, the school's librarian is also very enthusiastic about it, attributing increased collaboration between teachers and librarians to the adoption of the assessment.

Impact on Teachers

Windermere's three sixth-grade teachers all praise the Sixth Grade Research Performance Assessment. They all agree that the assessment is a valuable exercise for students in its own right, beyond its merits as an assessment tool. Thus, teachers and administrators assert that the assessment is "valid on the face of it — teachers validate it."

Each of the three teachers interviewed described how he or she had modified instruction in the classroom to emphasize the skills students need to develop for the research assessment. One of the teachers has modified the four scales used on the Sixth Grade Research Performance Assessment for use in her classroom. She tailors her classroom scales to each research project — for example, projects on space, rain forests, and American wars during the 1993-94 school year — she assigns her students.

Another teacher commented, "I use a lot of visuals to prepare them for the assessment. For example, we use M&M's to practice graphing skills in math. I emphasize visuals not just in social studies but in all subject areas."

Finally, the third sixth-grade teacher explained that he had held practice oral presentations throughout the school year — six to eight over the course of the year — and that one of these was presented at night before the students' parents; thus students gained experience giving oral presentations before an audience of adults, and parents gained familiarity with one portion of the assessment their children would undergo.

This testimony from all three of the school's sixth-grade teachers embodies the words of the Windermere librarian: "Teachers are teaching to the assessment, which is exactly what we want them to be doing." Both the librarian and the sixth-grade teachers also reported that the school's fourth- and fifth-grade teachers also now emphasize research skills more in their instruction than they did formerly. According to the librarian, teachers now collaborate with her much more than they did formerly, and, consequently, library skills are no longer taught as discrete skills but as skills that children will have to learn to apply in other areas.

Teachers and district and school administrators all identify the same greatest challenge teachers face with the assessment. Namely, teachers find it very difficult to refrain from helping their students as they complete assessment tasks. "The hardest thing is not being able to help my students in anyway," reported one teacher. It is for this reason that students' classroom teachers may not be present in the library while students work on the assessment or in the room where students deliver their oral presentations.

Finally, the librarian reports only one minor gripe from school staff with respect to the assessment. Because the students work on the assessment in the library, the library is closed to the rest of the school during the days on which the assessment is administered, resulting in a grumble or two.

Impact on Students

The Windermere sixth-grade teachers say that their students' research skills have improved over the three years since the research assessment was introduced; district administrators indicate that they have discerned the same trend. With the pilot assessment, the district discovered, to nobody's surprise administrators say, that students were not demonstrating the research skills the district's teachers said they valued. With the increased emphasis on teaching research skills has come increased student facility in the area.

Teachers say that their students typically like the research assessment. "This is more real to students [than multiple-choice standardized tests]," said one teacher; according to another, "It works for them." Teachers do note, however, that some students find the assessment process to be very stressful.

Teachers also report that the assessment has revealed to them some abilities their students have that might otherwise have remained hidden. "Some kids have surprised us with their oral presentations — kids who aren't really strong writers who really shine for the presentation." Additionally, the assessment is flexible enough to accommodate various learning styles: "Some kids plan their visuals ahead of time . . . the visual learners often prepare their visuals before they write their reports."

Students demonstrate a reasonable understanding of the purpose of the research assessment, though the students who participated in this case study seemed to identify the oral presentation (only one of four equally weighted dimensions) as the most important aspect of the assessment. For instance, when asked what the purpose of the assessment was, one girl said, "How to speak in front of people, because we might need that skill later on in life." She also said that the assessment is a good means of achieving that objective. Another boy also identified the presentation as the purpose of the assessment — "It helps you get in front of people and present." He, too, said the assessment is a good way of achieving that purpose, though he said the presentation was the most difficult part of the assessment for him. These two students said that their teacher helped prepare them for the assessment by assigning two or three practice research projects with oral presentations. They said the practices were helpful, though they had more time to work on them than they had on the actual assessment.

Teachers also commented that the research assessment helps teach students organization and planning skills. Two comments include: "They need to know how to work under pressure, and they need to know how to research," and "We've been working toward the assessment all year [so they know what's expected of them]. The difficult part for them was the time pressure."

Impact on Special Education Students

One teacher commented that her classified students really struggle with the assessment. "It's overwhelming for them . . . . A *12* is passing, but if my classified kids get *10's*, I'm thrilled." A parent of a classified student echoed this sentiment. She was happy her daughter had participated in the assessment but felt that the girl would feel more of a sense of accomplishment if the assessment were administered or scored differentially for her. Although this parent approves of the concept of the assessment and believes that learning research skills is valuable, she objects to the scoring process: "It is detrimental to have strangers come in and hurt the children's self-esteem."

Impact on Parents

Teachers report that, for the most part, parents have responded positively to the assessment. Parents' biggest concern, according to both the sixth-grade teachers and the one parent interviewed for this study, is that, as it exists now, the assessment generates a good deal of stress for students but to no clear end. The assessment is not currently used for purposes of placement in junior high school (placement conferences are held early in the spring, prior to the administration of the assessment). Nor has the junior high school developed remediation specifically tailored for students whose performances on the research assessment were subpar. The interviewed parent — and, indeed, the district as well — would like to see the Sixth Grade Research Performance Assessment tied into the district's curriculum in a more comprehensive way.

Future Plans

There is discussion in the district of developing a research assessment for eighth-grade students as well; by doing so the district would address the concerns of those who feel as the aforementioned parent, i.e., that the research assessment is currently a one-time event not connected coherently to what students experience during the next stage of their schooling. However, the district has no firm plans to develop an eighth-grade assessment at this time.

Neither the Assistant Superintendent for Curriculum and Instruction nor the Windermere teachers and librarian anticipates major changes to the assessment in the near future. They did, however, identify aspects of the assessment that may be modified:

- The trait scale for scoring students' visuals may be modified. According to the librarian, "We're still not happy with it. We tried to make some changes this year but couldn't come to any agreement, so we left it as is."

- The district may undertake efforts to improve the standardization of administration of the assessment across schools. Not only is the assessment conducted over either two or four days at different schools, but, according to teachers and the librarian, students at different schools often have different levels of help available to them. In addition, library and other building resources vary across schools, perhaps a more difficult obstacle to standardization. If the research assessment becomes a "high-stakes" event in the future (e.g., if it becomes tied in with junior high school placement), standardization might become a more important issue for South Brunswick than it is now.

- The training provided to assessors will continue to be improved. Specifically, the district has plans to expand the amount of training assessors receive for scoring visuals.

Conclusions

The South Brunswick Sixth Grade Research Performance Assessment seems to be a valuable educational experience for students finishing their elementary school years. Teachers and students alike respond positively to the assessment.

As it exists now, however, the assessment is a one-time, low-stakes experience with apparently little or no impact on students' future educational experiences and only minimal feedback for instruction. Because the assessment takes place at the end of the sixth-grade year, teachers can use the information obtained from the assessment only to modify the instructional practices they will employ with their *next* group of students. Any particular insights the assessment may reveal with respect to *current* sixth-grade students can have no positive consequences in the classroom for the students who are about to leave elementary school. Nonetheless, the feedback teachers receive from the assessment about the effectiveness of their instructional techniques is likely to enhance learning in South Brunswick's schools over the long run.

Graduating sixth-graders move on to junior high school, where their teachers may or may not be aware of their performance on the assessment and, hence, of the students' strengths and weaknesses in research skills. However, the district has plans to develop an eighth-grade research assessment, thereby introducing greater coordination in the teaching and assessing of research skills across elementary and junior high school. If this plan is carried out, the Sixth Grade Research Performance Assessment in the future will be an integrated facet of a more comprehensive system of performance assessment. Right now it represents a strong first step.

APPENDIX A

Elementary Expectations

HOLISTIC STATEMENTS ABOUT IMPORTANT THINGS SIXTH GRADERS SHOULD BE ABLE TO DO WHEN THEY LEAVE THE ELEMENTARY SCHOOLS

When students leave grade 6, they should:

- ► Use information and concepts studied in Social Studies and Science during the elementary years (■ #)
- ► Recall core information set forth in the various curricula
- ► Identify, express and solve problems (math, verbal and visual) (• +)
- ► Enjoy reading, writing and problem solving
- ► Take risks and experiment with language, numbers and visuals (■ •)
- ► Ask a question that is appropriate for research (■)
- ► Identify the type of information needed for a task, locate that information and organize it appropriately (■ •)
- ► Present information and ideas clearly through written, visual and oral means, demonstrating awareness of audience (■ # • +)
- ► Present ideas through one or more of the following: art, music, dance, writing, construction (#)
- ► Make choices among the methods of expression and know which are most appropriate for them and for the task (# +)
- ► Explain the processes and procedures (strategies) they use and explain why they were chosen (# • +)
- ► Work successfully alone (■ •)
- ► Work successfully as part of a group (# +)
- ► Persevere at a difficult or complex task (■ # • +)
- ► Take ownership of their work and take pride in it (■ # +)
- ► Show initiative (■ # • +)

■ Research Assessment, February, 1991
# Arts Assessment, June 1991
• Math Assessment, June 1991
+ Problem-solving Assessment

APPENDIX B

Performance Criteria Assessment

CRITERIA FOR PERFORMANCE ASSESSMENT — SOUTH BRUNSWICK TOWNSHIP PUBLIC

Name of Student ________________ Name of School ________________

Question: ____

Written Scale

Please indicate the number which best describes the overall quality of the written work. If the student can do everything at one point but not at the next, mark 5. (For example, there could be a score of 3.5.)

5. Complete, well written, elaborate, convincing
   - ► The student clearly responds to the question and elaborates on the question as well as its importance. Details and examples are given to fully support the answer in question.
   - ► An introductory paragraph clearly shapes the question and the concluding paragraph has clarity and insight.
   - ► The writing is engaging, organized and fluid.
   - ► Sentence structure is varied and mechanics are correct. (No obtrusive errors.)
   - ► Varied sources are organized alphabetically, including author (if known) title and publication information.

4. Complete, convincing, conclusive, competently written
   - ► The student addresses question through introductory paragraph.
   - ► The student sufficiently answers question and provides some supporting details.
   - ► The writing is interesting and organized.
   - ► Structure and mechanics are consistently correct.
   - ► Concluding paragraph is included and varied sources are cited in an organized way.

3. Basically competent: satisfied requirements
   - ► The student states the question.
   - ► He/she answers the question with a small amount of supporting information. There is a basic organization but it is not always clear.
   - ► The structure and mechanics are generally correct with some errors.
   - ► Two different types of sources are noted.

2. Inconclusive or unclear
   - ► The student has a question but the answers and/or conclusions are undeveloped or irrelevant. Basic information may be lacking.
   - ► The writing may lack organization and be difficult to follow.
   - ► There may be many errors of sentence structure and mechanics. Sources may be mentioned.

1. Incomplete, incoherent, poorly written
    - ▶ The student does not state the question. No answer or conclusion is given.
    - ▶ Writing is disorganized and difficult to read.
    - ▶ Sentence structure and mechanics are consistently weak. Sources may or may not be noted.

TOTAL WRITTEN SCORE ______

Assessor's Comments:

Name of Student ____________________ Name of School ____________________

Question: ____

Trait Scales

Circle the number that best reflects the presentation. Add the numbers and provide a total score.

1	1/2	0
The presentation totally fulfills all elements of the criteria as stated.	The presentation only partially fulfills the criteria as stated.	The presentation does not fulfill the criteria as stated.

Visual		Oral	
The visual presentation:		In the oral presentation, the student:	
Is neat and easily visible	1 1/2 0	Has a clear and interesting introduction to the topic	1 1/2 0
Shows student creativity	1 1/2 0	speaks clearly, audibly and with inflection	1 1/2 0
Appropriate and engaging to its audience	1 1/2 0	Has eye contact	1 1/2 0
clearly enhances the question studied	1 1/2 0	Uses verbal strategies that engage the audience (such as metaphors, rhetorical questions, colorful examples, strong verbs)	1 1/2 0
Is a well integrated part of the overall presentation	1 1/2 0	Gives relevant supporting data to convincingly answer the question	1 1/2 0
TOTAL VISUAL ____		**TOTAL ORAL** ____	

Name of Student ____________________ Name of School ____________________

Question: ____

Process Scale

The student shows evidence of a plan designed to achieve written, oral and visual goals.	1 1/2 0
The student's plan is appropriate to accomplish the task.	1 1/2 0
The student demonstrates that he/she can use available resources to locate appropriate materials.	1 1/2 0
The student shows evidence of using writing process: note taking, drafting, revising, editing.	1 1/2 0
The student can critique, reflect on, or analyze the progress of his/her work.	1 1/2 0

TOTAL PROCESS ____

TOTAL SCORE ____

Names of Assessors:

New York Regents Portfolios:
Hudson High School
May 12-13, 1994

HUDSON

NEW YORK REGENTS PORTFOLIOS: HUDSON HIGH SCHOOL

Introduction

This case study examines a performance assessment system designed by an individual school in Hudson, New York in response to impending changes in and opportunities provided by New York's state testing system. The subject of the study, Hudson High School, is a suburban school for grades 9 through 12 adjacent to the cities of Albany and Schenectady. It is Hudson Central School District's only high school. Hudson's student population ranks at the state median[1] on measures of wealth; only 2.5 percent qualify for free or reduced lunch. Of Hudson's 1393 students, 91.1 percent are Caucasian, 3.3 percent are Asian, 2.9 percent are African-American, 1.8 percent are Hispanic, and 0.9 are classified as "other." Last year, 83 percent of the school's student body went on to some form of postsecondary education.

Thirty-six of the 103 Hudson High School teachers are formally trained in the use of performance assessments and are using them as full or partial replacements of the New York State Regents Examinations in English, biology, earth science, and social studies. Twenty-three others, although not formally trained, are using performance assessments in some way in their classrooms.

Participants

At Hudson, the people listed in Exhibit I were interviewed.

Observations

Observations include performance assessment oral presentations in biology, English 11, earth science, and a video of a 10th grade integrated English/social studies dramatic production.

State Context

New York has the oldest and largest state testing program in the nation. Currently, 16 high school Regents Examinations are offered in the areas of English, social studies, mathematics, sciences, and foreign languages. The examinations are intended for the college-bound and are taken by approximately 60 percent of the state's high school population. They are used to maintain state-set educational standards, influence instruction, provide accountability (schools' test scores are reported to the public annually), demonstrate individual competency, and help make

---

[1] From Guilderland high school's application narrative to the state education department for a U.S. history/English 11 integrated course waiver of the New York state Regents Examination.

EXHIBIT I

Study Participants

- Principal, Hudson High School
- Sixteen teachers who are using performance assessments, including one foreign language teacher, four social studies teachers, five English teachers, four earth science teachers, and two biology teachers
- Five teachers who are not using performance assessments, including one social studies teacher, two mathematics teachers, one physics teacher, and one chemistry teacher (Two of these teachers also serve as teacher union representatives)
- Two 10th grade and two 11th grade students
- Three parents (one of whom is president of the local Parent Teacher Association (PTA)
- One school board member

college application and admissions decisions in the state of New York. Most of the Regents Examinations offer a mix of essay and multiple-choice questions with two main exceptions. The modern language examinations (e.g., French, Spanish) consist of an oral communication performance, and a written test of listening comprehension, reading comprehension, and writing skills. The science examinations (biology, chemistry, earth science, and physics) consist largely or entirely of multiple-choice questions.

In 1992, however, New York's Commissioner of Education and Board of Regents decided to overhaul the existing state testing program as part of New York's *New Compact for Learning* — a comprehensive strategy for improving public education in the 1990s. The redesign is intended to address some of the problems with the current state assessment system, such as the fact that it:

- duplicates local district or school-based testing programs;
- provides externally scored data on a delayed time frame that cannot support teaching decisions;
- encourages instruction that emphasizes rote-oriented tasks focused on lower-level knowledge and skills, rather than rigorous content and higher-level performance abilities;
- does not promote linkages with the world of work or ask students to demonstrate their proficiency as they would in real-life situations.

The new system envisioned and proposed by the New York State Curriculum and Assessment Council in November 1993 would move the State from a state-developed testing regime that focuses on summative evaluation using primarily multiple-choice forms of testing to a system of locally-designed curriculum-embedded performance assessments that will be tied directly to instruction and to the improvement of instructional practice.

Under the new system, all high school students would assemble a Regents Portfolio for Graduation growing out of a locally-designed, but state-approved K-12 student portfolio. The Regents Portfolio will include both discipline-specific and multidisciplinary student work samples (papers, projects, and exhibitions), and performance assessments that demonstrate competence across the seven curriculum areas and the state's set of Essential Skills and Dispositions. The current Regents Examinations would be replaced by a set of these performance assessments which could include on-demand or extended performance tasks, science and social studies research projects, writing and arts portfolios, or oral presentations and defenses of major work products.

Development of Performance Assessments

The history of Hudson High School's performance assessments is closely linked to the changes taking place in the state testing system. Several years ago, a global studies teacher and an English teacher at the school requested authority to develop a 10th grade multidisciplinary, team-taught course that made connections between history and literature and focused heavily on writing assignments and performance-oriented tasks. The course was reviewed by subject area departments, the school supervisory team, curriculum cabinets, and the district administrative team and approved by the district's Administrative Council. These teachers complained, however, that the Global Studies and English Regents Examinations were "strait-jackets" that prevented them from implementing their new course successfully. Therefore, they petitioned the State Education Department for a regulatory variance that would substitute a portfolio of demonstrations containing a reading response journal, a multiple source paper, a persuasive essay, and a biography project for the Regents Examinations in global studies and English. This variance request was well-received and approved by the state in 1992.

The following year, in 1993, a U.S. history and government teacher and an 11th grade English teacher developed a similar integrated course for the 11th grade which also received approval from the district and from the state for Regents Examination exemptions in English 11 and U.S. history and government. In 1993-94, the integrated course and examination waiver was extended to the ninth grade.

In July of 1993, after the first integrated courses with full Regents Examination waivers had been developed at Hudson, the State Education Department invited representatives from schools all over the state to a four-day conference to consider pilot development of performance assessments that would serve as *partial* (up to 35 percent) waivers to the Regents Examinations in the sciences, English, and the social studies. Teachers described this as an "unbelievable departure from tradition" that clearly indicated that Commissioner of Education Sobol was serious about assessment reform.

After the meeting, Hudson's principal persuaded the "leaders" in his science, English, and social studies departments to develop proposals to take advantage of the freedom the state's partial waiver offer allowed. As an incentive, he offered 5 days of paid staff development time for interested teachers. As a result, all earth science, biology, English 11, and Global Studies I and II (9th and 10th grade) teachers implemented a partial waiver curriculum in at least one of their class sections during the 1993-94 year. At a follow-up state meeting in September, the principal learned that Hudson was the only school in the state to attempt the waiver in more than one subject and one of only five schools to attempt it in any subject. Both the principal's leadership and the school's prior history with teacher-initiated courses that received full Regents waivers seem to have been instrumental in encouraging wider adoption of the state proposal at Hudson.

Assessment Characteristics

Hudson's teachers have developed performance assessments as full or partial waivers to the Regents examinations in: 10th and 11th grade English and social studies (through integrated courses), 11th grade English, 9th and 10th grade global studies, 9th grade earth science, and 10th grade biology.[2] In addition, a performance-based version of the Regents Examinations in Spanish and French has been employed by schools across the state since 1991. As described below, students have the opportunity to earn a predetermined number of points out of the 100 points on each of the subject area Regents Examinations by completing these performance assessments. The assessments are graded by classroom teachers, who also may assign them traditional letter grades that figure into a determination of overall course achievement.

Integrated English/Social Studies

In the integrated courses, literature and writing are used to enhance and enrich the Global Studies (10th grade) or U.S. history and government (11th grade) programs. Reciprocally, issues raised in the social studies provide subjects for in-depth writing and further reading. Because of the emphasis on depth of knowledge, less time is spent learning a wide breadth of facts and isolated concepts. The integrated course is designed to: (1) provide interconnectedness between the two disciplines and between past and current history, (2) improve students' analytical ability and research skills, (3) foster writing across the curriculum, and (4) encourage students to study different historical viewpoints and to draw their own independent conclusions.

In the 9th and 10th grade global studies/English courses, students must complete: (1) a "persuasive sequence" that includes an argumentative essay, a multi-source research paper, and a persuasive speech; (2) a bi-weekly journal responding to class discussions, readings, and presentations, (3) a "meeting of minds" project in which students adopt historical persona and produce a scripted and video-taped debate or drama, and (4) a historic literature project in which students read an important work and write a commentary on its historic and literary significance. To earn honors credit, students must complete two additional projects such as foreign "pen pal" correspondence, with international students, or a creative "time-travel" short story based on the life of a significant historical figure.

In the 11th grade U.S. History/English course, a portfolio of six projects is used to replace the Regents Examinations for English 11 and U.S. history and government. These include: (1) a weekly response journal, (2) a multi-source paper on African-American or Native American history, (3) a writing portfolio of 22 finished pieces in a variety of genres, (4) a biography project on a woman or minority of historical significance, (5) an anthology of literature on a particular historical subject or a family history project, and (6) an "American Dream" personal perspective paper that is a culmination of the entire year's thinking and learning. Samples of rubrics used to evaluate these projects are provided in Appendix A.

---

[2] Hereupon referred to as "waiver courses."

Eleventh Grade English

In place of the reading comprehension, spelling, and vocabulary sections of the Regents Examination in English (35 points out of a total 100), students in this course complete: (1) 12 reflective essays about works read and class activities and discussions (12 points), (2) three quarterly reading self-assessments discussing information and understanding gained from course readings and personal strengths and weaknesses in reading (3 points), and (3) a final reading project and oral presentation centered around a particular author or literary theme (20 points). A sample of the rubrics used to evaluate these projects are based on the New York State Outcomes for Reading and are provided in Appendix B. Students must still complete multiple choice listening comprehension questions (10 points), a literature essay (25 points), and a written composition (30 points) on the June Regents Examination.

Global Studies

Twenty points of Hudson's performance assessment component of the Global Studies Regents examination are completed in 9th grade, and 30 points are completed in 10th grade. In 9th grade, students complete a 10 point written research project and two five point projects that might include: writing assignments such as essays, poems, journals, and stories; a group research paper and oral presentation on a particular country's culture, history, and economy; or a creative project which could involve performing arts, speeches, political cartoons, maps, or newspapers. In 10th grade, students again receive 10 points for a second research paper, and 20 points for completing two of the additional projects described above.

Earth Science

In Earth Science, students earn up to 20 points on their Earth Science Regents Examination by completing a long-term study that requires understanding of key scientific and geologic concepts and promotes development of observational, analytic, and investigative skills. At the beginning of the year, students are presented with a "pet rock" of unknown composition and origin about which they must learn as much as possible throughout the year. They must keep a detailed scientific journal containing their observations, inferences, and predictions about the rock's scientific characteristics, genesis, metamorphosis, geographical location, relationship to the earth's structure and weather patterns, and commercial value.

In the second semester, they must "go beyond" the journal in their investigation of the rock's relationship to the environment and other scientific fields. For instance, they may develop a laboratory study to test the rock's composition, take a field trip to a site where the rock might be found, or interview a professional geologist about the rock. They must end their investigation with a multi-media oral presentation summarizing their year's research. Criteria by which Earth Science journals and oral presentations are judged are presented in Appendix C.

Biology

In biology, students complete three projects and a lab practical exam that substitute for 35 of the 100 points on the Biology Regents Examination. This year, the first project was an invertebrate research paper that included drawings of the chosen organism's structure and a written discussion of the organism's environment, life cycle, patterns of behavior, and environmental adaptations. The second project was a controlled experiment in which students worked in groups to study the effect of radiation on plants' growth. For the third project, students were asked to interview or "shadow" a life sciences professional and to report orally on the their experiences in class, using multimedia aids. In the lab practical exam, students had to demonstrate use of a microscope and identify structures of various plants and animals.

Foreign Languages

According to the Foreign Languages Supervisor at Hudson, the Regents examinations in French and Spanish have been performance-based since 1991 — although because of the nature of foreign language instruction assessment, they have always been at least somewhat performance-based. Students must complete writing, speaking, listening, and reading portfolios, and grades are determined by the number of "menu items" or projects they choose to complete for each portfolio. On the listening portfolio, students can choose to listen to and write summaries of French or Spanish television programs, music videos, plays, or films. For the speaking portfolio, they can conduct interviews in French or Spanish with native speakers, audiotape conversations in French or Spanish, or prepare an oral presentation on a cultural topic of their choice.

Assessment Quality and Consequences

No formal evaluations of the validity and reliability of Hudson's new performance assessments have been conducted as of 1994. Teachers at Hudson who are using the performance assessments *believe* that they provide a more realistic appraisal of students' abilities and therefore represent a more valid diagnostic tool than the Regents Examinations. Teachers noted that the content of two-year courses such as Global Studies is too broad for knowledge to be accurately assessed by a three-hour Regents Examination. Performance assessment projects, on the other hand, allow students to demonstrate what they do, rather than do not, know without artificial time constraints.

Hudson teachers were concerned about the reliability of their performance assessments, since they had not received any training in rubrics development or scoring. Most teachers had developed simple rubrics on their own; some of these are presented in the appendices. However, most were unsure about the consistency and equity of their grading methods, particularly those, such as science teachers, who were grading lengthy writing assignments for the first time.

Resource and Staff Development Support

Performance assessments also require more resources, materials, and space. Teachers note that portfolios and other physical evidence of performance assessments are bulky to store, and that performances by their very nature require greater space and more equipment. For instance, biology teachers at Hudson lacked sunlit space to conduct their seed experiments, while social studies teachers provided their own audiovisual equipment to tape students' historical dramatizations. Since students are given greater choice about the sources from which they learn, teachers also must provide a greater range of books and resource materials. One teacher at Hudson noted that "I am spending a fortune out of my own pocket to ensure students have appropriate resources."

To date, the staff development costs of performance assessments at Hudson have been quite low, but only because little training or paid planning time has been offered. Teachers who agreed to develop partial waivers to the Regents Examinations at the invitation of the state were paid for one or two days of group planning time last summer. However, this compensated for only a small portion of the total time they spent planning new curricula, assessment exercises, and rubrics. These teachers devote free time to performance assessments as a "labor of love," but few others said they were willing to be so charitable. Even the most dedicated teachers expressed frustration with their work loads and were uncertain whether they could sustain their level of effort, particularly if asked to teach more than one section of the waiver courses.

Teachers were divided about whether they needed specific training in the development and use of performance assessments. Those who had piloted their use seemed to be more creative and resourceful and less traditional than most teachers at Hudson; many had read assessment literature on their own, initiated brainstorming sessions with their colleagues, and attended university courses at their own expense. However, parents and administrators acknowledged that most teachers would need a great deal of guidance and support to teach this way. All teachers spoke of the need for common planning time with their departmental colleagues in which to develop testing approaches, revamp curricula, and ensure scoring consistency, particularly when expert training is not available.

Impact of Performance Assessment

The following section examines the specific impact of the Hudson performance assessments upon the major constituencies at the school.

Impact on Teachers and Administrators

Hudson's performance assessments impose a tremendous planning and grading burden on the teachers who use them. Although teachers who use them are largely enthusiastic about the benefits of the waiver courses for students, they feel the workload these courses create is "almost suicidal." Since most portfolio products involve extensive writing, they take much longer to evaluate than the standard multiple-choice or short-answer test. Currently, teachers have up to 90 written assignments to grade on an almost weekly basis, and district financial pressures threaten to increase class size at Hudson in the coming years. Several teachers estimated that one section of a waiver course creates up to three hours of additional work each day. Although planning will become easier as teachers gain experience with alternative assessments, the grading load is not likely to lighten.

Hudson's principal recognizes the additional costs of the performance assessments developed in waiver courses. Since waiver courses are often team-taught, require smaller class sizes, and increase teachers' work loads, they also increase school costs. To date, teachers have not received additional compensation for teaching waiver courses, but union contracts will have to be renegotiated as the courses gain in popularity and increasing numbers of teachers are asked to teach them. As Hudson's school board member acknowledged, "we can't afford to do this, and we can't afford not to either."

Performance assessments also require more teacher creativity, initiative, and independence. Some teachers at Hudson, particularly those who have taught in New York for many years, were uncomfortable relinquishing the structure and familiarity of the Regents-mandated curricula and examinations. Teaching to the Regents tests assured them that they were covering appropriate material; one teacher said he "trusted the state to tell me what students need to know." The demise of the Regents Examination system would bother them because they believe the exams safeguard minimum levels of learning and a "standard core of basic knowledge." Regents Examinations also minimized their planning time and the need to take risks, such as sending students out into the professional community or staging a night of student-scripted work for parents.

Impact on Curriculum and Instruction

Performance assessments drive instructional change and improvement. Teachers said that the Regents waivers had given them the freedom to "let go" of the tight schedules and textbook-oriented assignments that had been dictated by the fact-based Regents Examinations. As a result, they were allowing more group work, exploration, oral presentations, debate, and creative projects to take place in their classrooms. They were learning to be creative in developing assignments that challenged and engaged their students, made interdisciplinary and "real-world" connections, and reinforced critical skills and concepts. One teacher remarked that "this has been the single best professional growth experience I have had in 27 years of teaching." In addition, unlike Regents Examinations, performance assessments provide immediate feedback about student progress.

Impact on Students

Teachers and parents felt that Hudson's performance assessments are more challenging intellectual exercises, requiring students to think, reason, and analyze at a much higher level. In social studies, for instance, portfolio writing assignments ask students to think critically about history's significance, rather than to regurgitate memorized facts. They must learn to integrate a broad range of historical and cultural perspectives and develop their own position on complex issues. Projects such as debates and "argumentative essays" teach them the value and challenge of developing a well-supported opinion. One student noted that facts tested on the Regents Examination are what she most easily learns and quickly forgets, while historical concepts she learned and opinions she formed with more difficulty through her portfolio assignments stick with her. Undoubtedly, she felt, her integrated English/social studies class had made her a "sharper thinker."

In earth science, the "pet rock" project requires students to develop their powers of observation, exploration, and analysis far better than a multiple-choice test. The project serves as a highly motivational tool to teach and reinforce key scientific concepts such as chemical structure, erosion, and deposition. Students are forced to puzzle through inconclusive evidence; as one teacher explained, they are "learning how to learn," reflect, and problem solve, rather than how to memorize large numbers of isolated facts. Learning is also more experiential and creative, mirroring the real world of professional science. As one teacher noted, "science is performance."

Performance assessments strongly promote the development of both oral and written self-expression. Projects such as the English/social studies "persuasive sequence" and "meeting of the minds" performance or the Earth Science "pet rock" presentation emphasize creativity and independent thought. Teachers talked about how much students enjoyed the learning process when they were encouraged to express their own opinions and given the freedom to respond creatively to challenges. One student discussed how pleased she felt to learn that "my opinions really matter." Performance assessments also require students to practice their writing and oral communications skills to a far greater extent than do Regents examinations.

Students also gain experience with a wide variety of kinds of writing through their portfolios and projects. For example, the biology invertebrate research paper assignment was the first experience students had with formal science writing, while in social studies and English, they were asked to write scripts for plays and mock trials, collaborative research papers, short stories, poems, and op-ed pieces. Students themselves emphasized how much writing they were asked to do in their Regents waiver courses and how much their writing skills had improved as a result. Surprisingly, they said they honestly enjoyed many of their writing assignments, because they were allowed to exercise creativity and to express their own opinions. They also felt that the courses had improved their research skills, since they frequently were asked to support and defend their positions.

Each one of the Regents waiver courses examined in the study asked students to develop oral presentations using multimedia or graphical aides. Many also asked students to contact and interview professionals in the community. Students felt these exercises had improved their verbal communication and persuasion skills, as well as their poise and self-confidence. As one biology teacher reasoned, "I'd rather have them learn how to make a professional presentation to their peers than memorize the details of the Krebs cycle."

Hudson teachers and parents felt that performance assessments improved the relevance of learning which, in turn, improved students' enthusiasm and motivation. Teachers felt that students learn more and learn better when they are having fun with their assignments and can see the relevance and significance of their work. Biology students felt that the projects in which they had worked with veterinarians, wildlife biologists, and medical personnel had raised their awareness of the way biology is used in the professional world. Teachers felt that slower students, in particular, benefited from assignments that clearly demonstrated "the concrete value of knowledge for everyday life."

Students said that assignments which asked them to adopt a historical persona or stage a mock trial gave them a sense of personal affinity and even emotional attachment to their subject matter. One mother said that her daughter's integrated history/English course had "turned her into a social activist." Teachers felt privileged to learn more about each students' personal opinions and perspectives and excited about having created a more collegial "community of learners".

Unfortunately, Hudson's performance assessments also occupy more class time than traditional tests. Teachers must often sacrifice portions of their curriculum in order to provide room for the discussions, role-playing, oral reports, and lengthy experimentation that performance assessments entail. Teachers at Hudson who use performance assessments have reconciled themselves to this situation because they believe that the depth and variety of learning opportunities that performance assessments provide are more important than mastery of broad content. However, several teachers we spoke with who do not use performance assessments — particularly those who teach higher level mathematics and science courses — feel that performance assessments would impede their ability to cover the breadth of material that is essential in their minds for preparing students for higher education.

Impact on Parents

Parents included in the study were unanimous in agreeing that performance assessments gave them more information about their children's progress. They noted that parents receive Regents Examinations scores months after their students take the tests and that the scores provide no information about their students strengths and weaknesses. On the other hand, they were more likely to see portfolio products, which allowed them to track their child's progress.

Future Plans

The future of Hudson's performance assessment program is tied closely to the future of the Regents examination system. The pilot program has worked well at Hudson, because it allowed both teachers and students to volunteer for the experimental program. However, the administration is likely to encounter a great deal of resistance should performance assessments be imposed universally. A union representative pointed out that "part of teachers' current love affair with performance assessment is the freedom it gives them from the state-mandated Regents system." He contends that teacher enthusiasm would quickly diminish if one mandated system was substituted for another, since he believes performance assessments are not appropriate for all teaching styles and all academic disciplines. District administrators agree that teachers must have "bought in" philosophically for performance assessments to work.

Performance assessments at Hudson have also benefited from strong school leadership and from a supportive school board that tolerates experimentation and risk-taking. However, despite strong moral support from school and district leaders, budget constraints that increase class sizes and minimize teacher pay and staff development opportunities threaten the continued success of Hudson's performance assessment movement.

Nonetheless, over the last 3 years, more than 50 high schools have followed Hudson's lead in working with the New York State Education Department to develop and implement partial waivers to the Regents Examinations in the sciences, English, and the social studies. However, these locally developed efforts are unlikely to continue to expand indefinitely. Instead, the New York State Department of Education has chosen to transition its resources into the research and development of a new state-wide performance-based testing system being piloted in schools in the Spring of 1995. While the initial pilots are quite modest in scope and are not replacing current assessments, they will eventually continue to expand until they can be implemented state-wide. Importantly, though, the locally developed performance-based assessments are serving to inform the development of this new, state-wide system. The input that Hudson's teachers and administrators are proving to be instrumental to the efforts of the state.

Conclusions

Hudson High School's first performance assessments were initiated by a few enterprising teachers as an integral part of an effort to increase the rigor and creativity of their curricula. These teachers believed that the state's traditional testing program was an impediment to more flexible and challenging instruction and successfully sought relief from the testing mandate. When the state later looked for ways to test the viability of performance assessments, Hudson was well prepared to capitalize on the state's proposals to encourage local experimentation.

The performance assessments that teachers at Hudson have designed as substitutes for the state Regent Examinations — and the creative curricular changes that have accompanied them — are extremely popular with the Hudson teachers and students who use them. All constituencies agree that the assessments demand more thoughtful, creative work from students and increase the relevance of academic subjects. Students are improving their writing, research, and debate skills like never before, while at the same time, enjoying their studies more than formerly. Parents feel more informed about their students' progress and more aware of their strengths and weaknesses. Some teachers who teach mathematics and science courses, however, feel that performance assessments would impede their ability to cover the range of material college-bound students need to know.

Unfortunately, however, the new performance assessments clearly represent extra costs in terms of resources and teacher workload. These costs will have to be shouldered by the school district if additional teachers and schools are to adopt the new approach. Steps also must be taken to ensure the reliability and validity of each assessment and to train teachers in the development of sound testing and scoring approaches. Successful resolution of these issues will likely bear upon the plans the state has for the development and implementation of the new testing program it is designing.

APPENDIX A
History/English Rubric

REGENTS EXAM PORTFOLIO RUBRIC: Multiple Source Papers

18 points

ORGANIZATION AND DEVELOPMENT 4	Clear Good Introduction Good Conclusion 4	Mostly Clear Introduction and Conclusion 3	Somewnat Clear Introduction or Conclusion 2	Unclear Lacking Introductic or Conclusion 1
DOCUMENTATION 3	Source List and Footnotes Complete and Accurate 3	Source List Correct Some Footnotes Omitted 2	Source List Somewhat Correct Most Footnotes Omitted 1	Errors in Source List Footnotes Omitted -5
THOROUGHNESS OF TREATMENT 3	All important information 3	Most important information 2	Some important information 1	Little important information -5
RESEARCH QUALITY INCLUDING NOTES 3	Variety of sources Notes complete 3	Some variety of sources Notes mostly complete 2	Little variety of sources Some notes 1	Few sources No notes -5
SKILLS AND MECHANICS 2	Completely accurate 2	Spelling accurate Few errors in usage 1.5	Most spelling and usage correct 1	Many spelling and usage errors -5
EFFECTIVENESS 3	Compelling 3	Very effective 2	Somewhat effective 1	Needs to be revised to be effective 0

Name____________________

Title/Subject #1____________________

Title/Subject #2____________________

Total points__________

Legge/Milligan
Integrated US History, English 11 (11X)
Regents Exam Portfolio Evaluation Sheet

Name ______________________
Date ________________

(35) Journals __________

(20) Writing Portfolio __________

(15) Individual Anthology Project __________

(18) Multiple-Source Papers __________

(12) American Dream Paper __________

Final Exam Grade __________

APPENDIX B
Regents Reading Comprehension Rubric

REGENTS READING COMPREHENSION

N.Y.S. Outcomes for Reading

These are the capabilities we hope will be fostered by our new emphasis on reading comprehension:

1. Aesthetic response: Appreciate and/or enjoy texts, relate texts to self, and respond sensitively to texts with diverse social, historical and cultural dimensions.

2. Information and understanding: Collect data, facts, or ideas, discover relationships, concepts or generalizations and use knowledge from text.

3. Critical analysis and evaluation: Use personal and/or objective criteria to form opinions or to make judgments about ideas and information in written texts.

Point Allocation

Reflection papers and the final project will replace 35 points on Part I of the English Regents. Specifically,

- 12 points will be given if 12 reflection pieces are written and earn a score of at least a 70.

- 3 points (1 point for each) will be given for a self-assessment paper written in class at the end of each of the first three quarters (one point per self-assessment).

- 20 points will be given to final projects:
  - up to 5 points for evidence of aesthetic response
  - up to 5 points for evidence of information and understanding
  - up to 5 points for evidence of critical analysis and evaluation
  - up to 5 points for presentation to include:
    - organization
    - awareness of audience
    - originality
    - cohesiveness
    - length - 10-15 minute presentations which may be a prepared reading of your paper or some other form of demonstration which will be accompanied by a written rationale
    - evidence of vocabulary development and accurate spelling

APPENDIX C
Earth Science Rubric

Criteria by which Earth Science Journals will be graded

A journal that earns *five points* will:

A. clearly describe questions studied and supply reasons for their importance.

B. clearly state conclusions in a thoughtful manner.

C. give a variety of facts, details and examples that demonstrate a skill or an understanding of a concept including correct labeling of units, graphs, tables, etc.

D. list all sources of information including personal observation and inference.

E. as the journal entries progress, show continual growth in the quality of answers to questions and resulting conclusions.

A journal that earns *four points* will:

A. adequately describe.

B. states conclusion in a thoughtful manner but with less clarity than a five journal.

C. gives facts, details and examples that demonstrate a skill or an understanding of a concept including most labeling of units, graphs, tables etc. done correctly.

D. lists most sources of information including personal observation and inference.

E. as the journal entries progress, shows growth in the quality of answers to questions and resulting conclusions.

A journal that earns *three points* will:

A. sometimes describe questions studied and supply reasons for their importance.

Criteria by which Earth Science Journals will be graded

A journal that earns *three points* will:

A. clearly describe questions studied and supply reasons for their importance.

B. clearly state conclusions in a thoughtful manner.

C. give a variety of facts, details and examples that demonstrate a skill or an understanding of a concept including correct labeling of units, graphs, tables, etc.

D. list all sources of information including personal observation and inference.

E. as the journal entries progress, show growth in the quality of answers to questions and resulting conclusions.

A journal that earns *two points* will:

A. sometimes describe questions studied and supply reasons for their importance.

B. sometimes state conclusions in a thoughtful manner.

C. sometimes give a variety of facts, details and examples that demonstrate a skill or an understanding of a concept including correct labeling of units, graphs, tables, etc.

D. sometimes list all sources of information including personal observation and inference.

E. as the journal entries progress, shows some growth in the quality of answers to questions and resulting conclusions.

B. sometimes state conclusions in a thoughtful manner.

C. sometimes give a variety of facts, details and examples that demonstrate a skill or an understanding of a concept including correct labeling of units, graphs, tables, etc.

D. sometimes list all sources of information including personal observation and inference.

E. as the journal entries progress, shows some growth in the quality of answers to questions and resulting conclusions.

A journal that earns *two points* will:

A. state the question but fail to fully describe it.

B. conclusions are vague or insufficient.

C. often lacks facts, details and examples that demonstrate a skill or an understanding of a concept.

D. sources of information may or may not be mentioned.

E. as the journal entries progress, shows erratic growth in the quality of answers to questions and resulting conclusions.

A journal that earns *one point* will:

A. rarely describe questions studied and supply reasons for their importance.

B. rarely state conclusions in a thoughtful manner.

C. rarely gives a variety of facts, details and examples that demonstrate a skill or an understanding of a concept including correct labeling of units, graphs, tables, etc.

D. rarely list all sources of information including personal observation and inference.

E. as the journal entries progress, shows little growth in the quality of answers to questions and resulting conclusions.

Maryland Assessment Reform:
Walters Middle School
May 9-10, 1994

MARYLAND ASSESSMENT REFORM:
WALTERS MIDDLE SCHOOL

Introduction

This case study focuses on the development of the performance assessment system designed by the State of Maryland and the impact and implementation of the system on Walters Middle School in Walters, Maryland. Walters Middle School is located in a suburb of the city of Walters, a "bedroom" community for the larger cities of Baltimore and Washington, D.C. Most families in the area are of middle or upper socioeconomic status; the median family income is $51,182 (in 1989 dollars). The community is predominantly white, and at Walters Middle School 96 percent of the 866 students are white, 2 percent are African-American, 2 percent are Hispanic, and 1 percent are Asian. Only one-half of 1 percent of Walters students receive Limited English Proficiency (LEP) services, and only 13 percent receive free or reduced-price meals. No students at the school receive Chapter 1 services. The school employs 55 full-time teachers.

The case study is based on state, district, and school documents, on interviews with a variety of people, and observations.

Participants

The roles of the people interviewed for this study are below in Exhibit 1.

Observations

Observations include students in five classes at Walters Middle School taking the MSPAP. (The MSPAP lasted for approximately two hours.)

State Context

Since 1991, the state of Maryland has been redesigning its entire accountability and assessment program as a part of Maryland's "Schools for Success" reform initiative. Deriving its ideas from the national education reform movement, the Governor's Commission on School Performance made the recommendation that such changes be undertaken.

The Maryland School Performance Program (MSPP), a system of monitoring how well schools and school systems are educating students throughout the state, includes data collected on student performance as well as supporting data on student population characteristics and financial information. One component of the data on student performance is derived through the new Maryland School Performance Assessment Program (MSPAP).

EXHIBIT I

Study Participants

- State Director of Student Assessment, Maryland State Department of Education;
- The Walters County Public Schools Associate Superintendent of Curriculum and Instructional Services;
- The Walters County Public Schools Director of Criterion-Referenced Evaluation and Testing;
- Two school board members;
- Walters Middle School Principal;
- Six teachers, including a special education teacher, two mathematics teachers, one science teacher, one language arts teacher (Department Chair), and one social studies teacher, all of whom are using both district and state performance assessments;
- Two additional teachers, including the school's teacher's union representative, and a *Renzuli Enrichment* teacher in charge of a special gifted and talented program;
- Several 8th grade students; and
- One parent.

Maryland School Performance Assessment Program (MSPAP) Characteristics and Development

In a departure from past practice, the Maryland Department of Education (MDE) is designing and implementing criterion-referenced performance assessments to replace norm-referenced tests as the primary vehicle for school and district accountability. The new Maryland School Performance Assessment Program (MSPAP) is intended to reflect standards of achievement commensurate with 21st century expectations. It also is intended to drive instructional changes that will help students to become better problem solvers and to learn how to apply knowledge to real world situations.

According to the State Director of Assessment, Maryland administered basic skills and minimal competency tests throughout the 1970s and early 1980s, and some of those tests "had a profound impact on what was taught." In some cases the focus of instructional activity became the tests themselves and resulted in a "narrowing" of the curriculum. For example, teachers in algebra classes focused on teaching arithmetic in preparation for the tests.

Beginning with the first administration in 1991, each May the MSPAP is administered to approximately 150,000 students in grades 3, 5, and 8, and all students in those grades ("the accountability grades") are required to participate in the assessments.[1] Each year, reading, writing, language usage, mathematics, science, and social studies are assessed, focusing on learning outcomes (both content and process) that were adopted by the Maryland State Board of Education in 1990. Each student in all of the grades assessed participates in 9 hours of MSPAP testing spread over a 5-day period at roughly 1 hour and 45 minutes of testing time per day.[2] (The student normally completes 2-to-3 assessment tasks per day.) Within each school, students and teachers are randomly assigned to classrooms for MSPAP administration; and classroom teachers are provided with training and with an *Examiner's Guide* to help them administer the assessments. The assessments are then scored in the summer by 600 scorers.

---

[1] The high school component of MSPAP is still under development.

Based on their scores on these MSPAP tasks, students are assigned to one of five proficiency levels in the assessed learning outcome: 1; 2, excellent; 3, satisfactory; 4; and 5. (Scores 4 and 5 imply that the work is less than satisfactory, but no official label is attached to them.) The score ranges for each of the proficiency levels within each content area and grade level are established for each MSPAP administration.[3] MSPAP scores range, in general, from 350 to 700 on each of the assessed areas.

MSPAP Tasks

Usually, MSPAP tasks require students to respond to a series of questions and other directions that lead to a culminating activity, such as a mathematical solution or a recommendation or decision. The tasks also require students to provide an explanation or rationale for their answers. The tasks are thus designed to tap into both the process and the content of thinking. Students are sometimes asked to work in small groups to understand the task and to collect data prior to recording their responses independently.

The MSPAP administered tasks are confidential and, therefore, are not available for public review. The Maryland State Department of Education, however, does share "public release tasks" with the general public, and one such 8th grade task is shown in Appendix A.

The "public release tasks" appeared in the 1992 MSPAP. They parallel the format of other MSPAP tasks. A 8th grade mathematics task called "Birth Dates," for example, provides students with information on the percentage of people born in each month of the year. Students are asked to graph this information. Next, based on the data given, they are asked to determine the number of their fellow students likely to share their birthday month. The entire task consists of four questions, or "activities," based on the data provided (see Appendix A).

MSPAP Task Scoring Rubrics

Each MSPAP task is accompanied by its own scoring rubric, and activities in each assessment task are rated on a scale of 0, 1; 0, 1, 2; or 0, 1, 2, 3, with the highest number indicating the most complete answer. The scoring rubric provides an overview of the type of competency or skill each activity elicits and explicates the criteria for assigning a score for the answer to the activity (given in *answer cues*). In addition, the rubric contains examples of student answers for each activity at each point on the scale. (See Appendix A for an example of a scoring rubric.)

---

[2] *Maryland School Performance Assessment Program*. Technical Report. (1994). Maryland State Department of Education.

[3] See, for example, *Score Interpretation Guide. 1994 MSPAP and Beyond. Maryland School Performance Assessment Program*. (February, 1995). Maryland State Department of Education, Baltimore, MD.

MSPAP Development

When the MSPAP system was initially conceptualized, the Maryland State Department of Education (MSDE) sought advice from experts such as Grant Wiggins and Lorrie Shepard, and organizations and networks such as the National Council of Teachers in Mathematics (NCTM) and the American Association for Advancement in Science (AAAS). Maryland learning outcomes were based on NCTM and AAAS standards as well as National Assessment of Educational Progress (NAEP) outcomes.

Currently, MSPAP assessments are designed, developed, and implemented by the MSDE in collaboration with Maryland classroom teachers and other educators. Assessment tasks for the 1994 MSPAP were developed by approximately 100 teachers and other school staff who were recruited, screened, hired, and trained by MSDE.

Task writers developed draft tasks according to MSPAP task specifications. Developers of scoring rubrics worked with task writers to determine *draft score point criteria* and *answer cue* information for each assessment.

Prior to administration, a task review is conducted by MSDE content and scoring specialists and staff from Maryland's testing contractor, Measurement Incorporated. The review is conducted to ensure that each activity will generate scorable responses that demonstrate evidence of the measured outcome. The assessment task review process involves reviews of content and clarity, reliability and validity, performance task characteristics, controversial and sensitive topics, and bias.

In November of the year preceding the May test administration, all MSPAP tasks are field tested in an out-of-state school district with a student population and educational practices closely matching those of Maryland. (Schools in districts such as Virginia Beach, VA and Pittsburgh, PA participate through CTB/McGraw Hill.) In December, about 45 trained Maryland educators and MSDE content and scoring specialists meet to score field test responses. Their job is to determine whether each MSPAP activity yields responses that reflect the intended outcomes (test validity) and whether the scoring rubrics facilitate efficient and accurate scoring. When problems are identified, the team recommends revisions to MSDE. In addition, team members identify possible "anchor student" responses, or "range-finders," for all possible rubric scores. Finally, responses are selected for use in scoring guides, teacher training sets, and teacher qualifying sets. To complete the training and qualifying sets, actual Maryland student response booklets also are "hijacked," or temporarily removed, immediately following the May assessment administration.

MSPAP Assessment Quality and Consequences

MSDE has instituted several measures to ensure the reliability and validity of MSPAP. These measures are briefly discussed below.

MSPAP Scoring

All scoring is performed by certified Maryland teachers (about 600 of them) at four regional sites in June and July, under the direction of MSDE and Measurement Incorporated. All participants receive on-site training from Measurement Incorporated. During that training, teams review scoring guides containing: (1) a description of the outcomes measured by the task (e.g., Reading for Literacy Experience, Writing to Express Personal Ideas, or Measurement with Estimation); (2) the criteria for awarding various amounts of credit (the scoring rubric is called a *scoring tool*); (3) *answer cue* information which includes possible text information or important ideas that might be included in a successful response); and (4) a set of sample student responses (anchor papers) to illustrate each of the score points that can be awarded.

To qualify as summer scorers, teachers are required to score with at least 70 percent exact agreement on at least one of three sets of training materials. Scorer performance is monitored throughout the summer to ensure that the scoring standards are maintained through administration of a series of *check sets* and *accuracy sets*, as satisfactory maintenance of the standards is a condition for continued scoring. Scorers work in teams under the direction of Team Scoring Coordinators and Team Leaders. Each scoring team periodically uses the *check sets* to recalibrate scorers and the *accuracy sets* to determine consistency. Each set contains assessment tasks and student responses representing all possible score points. (The representative student responses are consensually chosen by a team of scorers.) These student responses are rescored by the scorers. Any reader who attains less than 70 percent accuracy on any accuracy set is retrained. For their efforts, scorers are paid $10.00 per hour.

MSPAP Validity

MSDE has collected evidence of the MSPAP's *content*, *face*, and *construct* validity. The Maryland learning outcomes, which form the basis for learning, instruction, and MSPAP assessment activities, are based on recently developed national curriculum standards and national assessments. For example, the reading outcomes are based on National Assessment of Educational Progress (NAEP) reading assessment objectives, and the mathematics outcomes are based on National Council of Teachers of Mathematics (NCTM) standards for curriculum and evaluation. Moreover, the assessment tasks are developed by content area and grade specialists.

Each task development team is given instructions on which outcomes to assess. Completed tasks are reviewed for grade appropriateness, although some teachers continue to report concerns about the difficulty and appropriateness of MSPAP tasks. The MSPAP employs "authentic" and "real-life" situations as assessment tasks, and reading selections are full-length published works, rather than cut-and-paste excerpts. Evidence of internal consistency by content area is assessed with each set, and items that function differently for different student groups are flagged to inform the development of subsequent tasks.

MSPAP Consequences

MSPAP assessments are intended to provide information relevant to assessing school performance and to guiding school improvement plans and activities. Information provided by MSPAP assessments is primarily focused on schools, although information about performance of individual students also is available from the assessments. Currently, a *Maryland School Program Report* is made available to every Maryland public school parent, containing MSPAP results delineated by county, school, race/ethnicity, and sex. (See Appendix B for State of Maryland disaggregated data.)

MSPAP standards were intentionally set high, with the initial expectation that few Maryland schools would be meeting those standards in the first years of administration. The Maryland School Performance Standards Committee and the Maryland School Performance Standards Council participated in the standard setting process. The Committee included representatives from the local school systems and from MSDE, and the Council included representatives from local education agencies, local boards of education, the teacher's union, business interests, students, and the state legislature. The two groups agreed on the following school performance standard: *For a given school to achieve satisfactory performance in a particular content area/grade level, 70 percent of students must achieve satisfactory performance (level 3 and above). Furthermore, to achieve excellent performance, a school must be at satisfactory and 25 percent of students must achieve excellent performance.*[4] All schools are expected to reach the satisfactory standards by year 2000.

While preliminary testing began in 1991, the 1993 results will serve as the baseline for school accountability reviews. At the time of this study, in 1994, no school met the satisfactory standard in all subject areas on the 1993 MSPAP. Twenty-nine schools met the standard in at least one of the 3rd grade subjects (i.e., mathematics, science, and social studies), and 71 schools met the standard in at least one of the 5th grade subjects (i.e., reading, mathematics, science, and social studies). (3rd grade reading and 8th grade science MSPAP results were anomalous;[5] 3rd grade results were lower than expected, and 8th grade results were higher. *For accountability purposes in 3rd grade reading and 8th grade science, the first year is 1994 instead of 1993.* [pg. 8])

Sanctions have not been imposed to date. At some time in the future, schools that exhibit substandard performances will enter into "receivership," under which they will be restructured by the state. However, Walters County Public Schools (WCPS) officials interviewed doubt whether MSDE has either the expertise or the resources to carry out such a plan.

A limiting factor in the utilization of the MSPAP has been the length of time required for processing of scores, and thus their delayed receipt by districts and by schools. Prior to this year, the results have not been available until after administration of the subsequent year's test, allowing no time for review and instructional adjustment. (The 1993-94 results are expected to be released in December, 1994). The primary instructional monitoring system in Walters County, therefore, continues to be student performance on the formative and summative components of the district's performance assessment battery (the Criterion Referenced Evaluation System (CRES) discussed in the next section).

Resource and Training Support

MSDE provides money to schools to purchase materials for the assessment tasks that require such materials (science experiments, for example). MSDE also suggests alternative materials, should specified materials not be readily available.

---

[4] *Score Interpretation Guide. 1994 MSPAP and Beyond. Maryland School Performance Assessment Program* (February, 1995). Maryland State Department of Education, Baltimore, MD.

[5] *Score Interpretation Guide. 1994 MSPAP and Beyond. Maryland School Performance Assessment Program* (February, 1995) Maryland State Department of Education, Baltimore, MD.

To implement the MSPAP, MSDE employs a "Train-the-Trainer" model. Twenty-four Testing Directors in the field train school-based Test Coordinators once a year. These Test Coordinators then train teachers in their own schools. The number of training sessions varies from school to school. MSDE does not provide any funds for training at the school level.

The State Director of Student Assessment feels that the training system needs to be improved. The assessment administration is complex, as students are organized into groups and must conduct extended experiments, and, a great deal of preparation is needed for these activities. Another important training aspect MSDE plans to address is Maryland teachers' knowledge of the Maryland learning outcomes. The Director of Student Assessment believes that teachers are not aware of the learning outcomes assessed by the MSPAP.

Walters County Reaction to the MSPAP

The Walters County Public Schools (WCPS) officials included in this study support the MSPAP and regard it as "extremely" complementary to WCPS goals. In addition, they are pleased that the state has involved them in formulating the assessment system. However, they harbor some concerns about the system's consequences, and about the logistics of its implementation. District testing officials feel that the MSPAP's use as an accountability measure may be premature, given that ". . . the state has yet to pull off a test that is fully comparable one year to the next."

However, neither the school board members nor the Walters principal and teachers who participated in the study are concerned about poor MSPAP performance or the state sanctions it might lead to, since ". . . WCPS' results on the MSPAP, while leaving much room for improvement, have been consistently among the best in the state."[6] In fact, district testing officials made a promise to teachers when the MSPAP was instituted that if they used the district's Criterion Referenced Evaluation System (CRES) faithfully, the MSPAP would "take care of itself," meaning teachers would not have to make any additional adjustments to their teaching or curriculum.

The two board members included in this study had differing reactions to the MSPAP. One school board member, who also chairs the Curriculum Committee, believes that "MSPAP is the right thing to do, but it needs to be improved." She is supportive of MSPAP and understands "the state's perspective" — its need for accountability, especially because the data obtained from the assessments will be helpful in evaluating equity issues in education. She also believes that the MSPAP is important for WCPS, as it provides a "credibility and validity check" for the district's own assessment system, CRES, since the instructional and curricular premises of the two systems are quite similar. In addition, she believes that the information provided by the MSPAP is useful for comparing WCPS with other school districts in Maryland.

On the other hand, the second school board member interviewed feels that the MSPAP is a "primitive tool" which has been imposed as an accountability mechanism before its authenticity and quality were rigorously proved. He has doubts about whether it is testing the right content, since, in his opinion, "state committees never agreed on what is worth knowing" before they developed the MSPAP.

---

[6] Between 37.6 and 55.7 percent of Walters County students scored at or above the satisfactory level for 1993, exceeding the state percentages for all grades and content areas. Walters Middle School scores, in turn, were above Walters County Public Schools' (WCPS) means in all four subject areas, and the middle school's mean scores were the best in the county in reading and mathematics.

In the second board member's opinion, the MSPAP lacks the CRES' virtue of "starting and ending in the classroom" (i.e., all teachers are involved in CRES development and all receive immediate feedback from its use). Before imposing a statewide examination, he says, MSDE should have determined whether state intervention would improve education at the local level (i.e., the school district level). Although he acknowledges that the MSPAP may be useful for districts that have not previously established appropriate standards, he feels it simply interferes with the quality of education already in place in Walters County.

The one area in which both board members express concern is the amount of instructional class time lost for time spent taking the MSPAP. Also, because MSPAP results are not available soon after the assessment is administered, the feedback cannot be as readily utilized for school improvement purposes (such as refining the curriculum and adjusting instructional practices).

District Context

Walters County is Maryland's largest county and is experiencing one of the state's highest growth rates. The school system adds new entrants at a rate of almost 16 percent a year, as compared with an overall state increase of 11 percent a year.

CRES Characteristics and Development

In 1988, concurrent with the development of the MSPAP, the Walters County Public Schools launched a process of continued school improvement, involving the district-wide restructuring of curriculum and assessment. The heart of this new improvement system is a three-tiered *essential curriculum model* that generates subject area goals, individual course objectives, and five interdisciplinary outcome "learner behaviors." These learner behaviors are at the top of the essential curriculum pyramid, and it is these toward which the entire Criterion Referenced Evaluation System (CRES) is geared (see Exhibit II).

According to the Associate Superintendent of Curriculum, the purposes of the new system are to affect instruction and to nudge students beyond retention of knowledge and to an "active use of knowledge."

At the outset of the district improvement process, it became clear that Maryland's existing norm-referenced multiple-choice tests could not adequately assess mastery of the essential curriculum or determine whether students are able to apply what they have learned to real life problems. Maryland functional tests set a minimum level of achievement for a Maryland diploma, but WCPS needed an assessment system that set new and more rigorous standards of performance for adult professional success in the twenty-first century.

Traditional tests tend to limit instruction to lower-level skills and to easy-to-measure knowledge, such as grammar and basic math. The intent of Walters County's new system was to reverse the equation: *to determine that which is important for students to be able to do and then*

EXHIBIT II

Learner Behaviors in the Essential Curriculum

- Effective communications skills
- Problem solving and critical thinking
- Social cooperation and self-discipline
- Responsible citizenship in the community and environment
- Lifelong learning

to devise a system to measure it. In addition, existing tests were primarily designed for sorting and selecting students, rather than for instructional feedback and adjustment. Walters County was looking for assessments that could serve as tools to adjust instruction in ways that would enable all students to master the essential curriculum.

Since 1988, therefore, Walters County has been designing and implementing an assessment system that requires students to respond to open-ended types of performance tasks that assess logic, reasoning, and comprehension. This new Criterion Referenced Evaluation System is intended to strike a balance between knowledge and its application by emphasizing both content and process skills in its performance tasks. The system consists of formative (on-going) assessments and end-of-the-year summative assessments.

Development of assessments began at every grade level with language arts and mathematics. Additional courses and disciplines were added each year, with a plan eventually to cover all required and elective subjects and levels, such as Physical Education and the Arts, by the 1994-95 school year. Baseline results were reported for the 1991-92 school year in language arts and mathematics, and for the 1992-93 year in science and social studies.

From the beginning, administrators, a representative group of volunteer classroom teachers, consultants (Dennie Wolfe, Grant Wiggins, Lauren Resnick, and Dan Resnick), and members of the Maryland Assessment Consortium were involved in formulating an operational system.

CRES Tasks and Scoring System

The CRES formative, or ongoing, assessments consist of various types of extended response tasks that provide students and their teachers with opportunities for adjustment of instruction and curriculum and for reteaching throughout the year prior to the final, or summative, assessment. Summative assessments similarly consist of performance tasks requiring extended responses. For example, a Grade 8 social studies task requires students to design a pamphlet about an ancient civilization that includes:

- A map indicating the geographical location of the chosen civilization;

- Descriptions and illustrations of at least three contributions from different aspects of the civilization; and

- Explanations of how each contribution has influenced later civilizations.

Two scoring rubrics are applied to the task: the Problem Solving rubric and the "Explanatory Domain" rubric. The "Problem Solving" rubric is a scale of 1 to 4, with 1 indicating nonproficient responses and 4 indicating exceptional problem-solving strategies. The Explanatory Domain rubric is also a scale of 1 to 4, with an explanation at each point of the response characteristics essential to obtain that score. (See Appendix C for greater detail.)

The use of CRES formative assessments is not mandated, but it is strongly encouraged. At least half of Walters Middle School teachers use between 1 and 20 performance tasks each year, while some do not use the assessments at all.

Formative assessments are developed by Walters County teachers in summer workshops, field-tested in pilot schools, and maintained in an assessment library in the media center of every school so that all teachers have access to them.

Summative assessments were first implemented in WCPS schools in 1992-93, a year after the formative assessments were introduced. Like the formative assessments, they were developed during summer workshops by WCPS teachers and field-tested in pilot schools prior to county-wide implementation. Summative assessments are required at every grade level, but individual departments in each school may decide when to administer them. With the exception of social studies, the CRES summatives are administered in May at Walters, and each of them can last anywhere between 20 minutes and five class periods.

CRES Scoring and Validity

As previously discussed, the CRES assessments use a four-point scoring rubric for each area of measured performance as follows:

- A score of 4 indicates exemplary work,
- A score of 3 is labeled "proficient,"
- A score of 2 is labeled "approaching proficiency", and
- A score of 1 is designated "evidence of attempt."

The goal is to have all students progress to the point of mastery, defined as a minimum score of 3.

Although CRES summative exams are scored by individual classroom teachers, samples of student work from a variety of county schools are selected and passed on to those teachers in order to enable them to "anchor" their evaluations to real student performance. To ensure consistency of scoring, a random sample of student responses is collected from schools following summative assessment scoring. District scoring review teams then meet to determine the degree of agreement between their own scores and those assigned by individual classroom teachers. (Samples are not large enough to judge overall scoring reliability at the school level, but problems in specific content areas at specific grade levels can be identified.)

The results of the yearly reviews are used by the district curriculum specialists to plan additional training and related staff development programs. The results from 1991-92 were indicative of excellent consistency for the initial implementation year; most disciplines approached a 90 percent match between individual teacher and scoring team rubric scores.

District officials believe that the CRES system has content validity, as it is closely aligned to the curriculum. Results from a teacher survey regarding the essential curriculum and the CRES system show that the majority of district teachers consider the curriculum to be covering important matters, but only a bare majority believe the assessments to be covering the important areas of the essential curriculum.

CRES Assessment Consequences

CRES scores are appended to each student's report card and are also aggregated by the school and reported county-wide. However, district officials recognize the psychometric fragility of the CRES tests and consider them "works in progress" that must be continuously refined. As a result, they do not intend to use them for school or teacher accountability purposes in the foreseeable future, but only as tools to (a) improve instruction, (b) help teachers measure whether the WCPS Essential Curriculum has been learned, and (c) provide helpful information to school improvement teams. For example, last year the district held a three-week summer "Super Camp" for those students identified as weak in their abilities to successfully master the Essential Curriculum.

Staff Development Support

A number of teachers interviewed at Walters volunteered that it is difficult to maintain objectivity in assigning CRES scores to students whose abilities and classroom performance are familiar to them. Teachers feel the problem is compounded by a lack of adequate training in the use of scoring rubrics, since they receive only one day of formal training in each subject area; and new teachers allegedly receive only an hour or so of training as part of their one day district orientation. However, department chairs and curriculum leaders in the school are available for help. The district is now trying to urge department chairs to encourage more collegial team scoring within individual schools, but little release time is allocated for such an endeavor.[7]

School principals and curriculum division staff receive monthly staff development programs on assessment, reteaching, and instructional adjustment. In a "trainer of trainers" model, all county teachers receive yearly staff development from their department chairs. In addition, teachers participate in the activities of the Maryland Assessment consortium[8] and the MSPAP development (see below). Some members of the teaching staff also participate voluntarily in paid summer work in which formative and summative tasks and items and scoring rubrics are developed.

---

[7] Teachers receive only three days of paid staff development time a year.

[8] Established in 1991 under the leadership of WCPS the Maryland Assessment Consortium is a collaboration of 24 school districts attempting to: (1) design and deliver high-quality staff development programs on performance assessment for member districts and (2) develop, field test, validate, and disseminate formative performance assessments for use by teachers as part of their instructional programs. Participants receive training in task development, pilot testing, and anchor paper development and then serve as key leaders in the training of other teachers in their home schools.

Coordination Among the Assessment Systems

One of the complaints most frequently voiced by Walters County teachers and school board members is that valuable instruction time is lost testing students on both the CRES and MSPAP, which are designed for essentially the same purposes.

To the district officials, though, the two assessments serve different purposes, despite similar structure and intent. First of all, the MSPAP currently is limited to just three grade levels, whereas students at every grade level take CRES summatives. More importantly, however, CRES tests are scored by classroom teachers, and individual results are available immediately. Individual student MSPAP scores will be reported beginning with their 1994 administration, but, to date, the scores have not been received by schools until at least a year after the tests were given. As a result, MSPAP assessments cannot be used for the purposes of instructional feedback, a purpose CRES tests take for granted.

On the positive side, MSPAP is being used by WCPS testing officials as an *external validation* of CRES scores. Since MSPAP tests are graded *blindly*, they are free of the bias with which CRES assessments might be scored. WCPS officials are therefore comparing individual district school results on the two assessment systems to ensure that the same relative performance trends apply to both. For instance, since Walters students score better than the district mean on each of the four MSPAP disciplines, they would be expected to score better than the district mean on the CRES in those four disciplines. Relative trends proved remarkably similar for most schools this year; the few schools that performed better on the CRES than on the MSPAP will be over-sampled next year, since this trend could indicate that teachers may be "effort-scoring" or may be under some sort of pressure to inflate their CRES assessment scores.

Nonetheless, WCPS testing officials felt there was some disadvantage to developing the CRES prior to the MSPAP, as some adjustments had to be made to the CRES once the MSPAP was introduced. Since the Maryland learning outcomes differed somewhat from WCPS's Essential Curriculum, district officials are still trying to achieve a closer correlation between the two assessments. For example, one board member mentioned that the science curriculum and assessments currently are being revised for better alignment with the MSPAP.

In addition to CRES and MSPAP, students at every middle and high school grade must still pass the Maryland state "functional" tests in reading, writing, and mathematics to assess basic competencies. (Functional tests in all areas, except writing, are multiple-choice.) Students in all high school grades (i.e., 9 through 12) must pass these tests in order to receive a high school diploma. The state has articulated the competencies that are required for passing these tests. (Ultimately, once an 11th grade MSPAP is developed, it may substitute for the functional tests as a requirement for graduation.)

WCPS students in the elementary grades also prepare portfolios that contain their best work collected throughout the year — and a portfolio system is being contemplated for the middle school grades.

Impact of Performance Assessment

This section discusses the impact of MSPAP and CRES on the Walters Middle School Community.

Impact on Teachers and Administrators

The school board members and the Walters teachers interviewed expressed a great deal of conviction that the CRES assessments in particular, and the MSPAP to a somewhat lesser degree, had substantially improved teaching practices in Walters County. They felt that in combination with WCPS' teacher-designed Essential Curriculum, the CRES had, as a board member expressed it, ". . . challenged teachers to think carefully about what they want students to know."

District administrators believe that "quite a few teachers" use the formative CRES assessments. No district-wide statistics, however, are available regarding either the number of teachers using the CRES formative assessments or the frequency with which those assessments are used.

Opinion among interviewees nearly unanimously expressed that CRES feedback helps motivate teachers to instruct their students for the sake of understanding and to encourage among a larger percentage of those students a mastery of the subject being taught. Toward that end, teachers said they were approaching their subject more "creatively." One teacher said that the CRES had finally convinced her that isolated facts are less important to teach than are the interdisciplinary skills of writing and reasoning. Teachers also found that they used more cooperative learning techniques and visual aids to improve the overall instruction of their students. Among those interviewed, the school board member noted that teachers were learning ". . . to bring their own critical thinking skills to bear on the teaching styles and methods they were familiar with."

On the "downside," the "cost" of the performance assessment system is measured by teachers in terms of the scoring time required by CRES assessments. While it is true that teachers "volunteer" to score the MSPAP at paid summer sessions, Walters County Public School (WCPS) teachers all must grade their students' CRES examinations during their planning periods or free time. District testing administrators acknowledged that because ". . . there is a great need to find some relief for teachers" from the labor-intensive nature of performance assessment scoring, they plan to add staff development days to the school calendar.

A large number of teachers also complained of the large amounts of time required to plan and to set up the MSPAP group experiment test components. This year, those problems were apparently alleviated somewhat by allowing classroom teachers to review the MSPAP a week before its administration, and by giving teachers one hour of release time to prepare the tests on the days they are to be administered.

District administrators say, finally, that the development of the CRES and Essential Curriculum, along with staff development, represents a sizeable WCPS investment that is paying large dividends in terms of improved instruction and increased learning. Instructional changes in response to CRES, thus, seem to be present, but the daily presence of MSPAP in the classroom, at this point, appears to be marginal.

Impact on Curriculum and Instruction

Although, CRES formative assessments are to be used in the classroom on an on-going basis, the most frequent criticism of performance at Walters Middle School was expressed as the amount of time the CRES, MSPAP, and functional tests collectively took away from classroom instruction.[9] One teacher suggested that CRES tasks be integrated across disciplines, so that test administration would occupy less curriculum time. However, as previously mentioned, CRES assessments are changing as a response to the new curricular frameworks, and instructional strategies are being guided by CRES themselves. MSPAP has had a less obvious impact on curriculum and instruction.

Impact on Students

Teachers and parents at Walters believe that both the MSPAP and CRES assessments are "more interesting" and "more relevant" to students than are standard testing formats, as both groups feel that the newer tests motivate students to do better work. Students also expressed enthusiasm for the two performance assessments — and particularly enjoyed the group work and experimentation as well as the opportunity and encouragement to express personal opinions.

Teachers and students agreed further that the "new tests" were far more "challenging" than minimum competency tests. On the MSPAP administration observed for this study, for example, students were asked to *conduct multi-step experiments*, *analyze the results*, and *explain the deductive reasoning* used, in writing. Students also were asked to (a) analyze primary source material such as Congressional testimony, editorials, political cartoons, or literature; (b) reconcile the authors' position with their own; and (c) consider the "real-world" implications of the positions they (the students) chose to defend. One teacher summarized a widely shared opinion that ". . . this is what testing is supposed to be like."

Impact on Special Education Students

Many teachers, however, were concerned that the CRES and MSPAP performance assessments presented too great a challenge for special education students and students with learning disabilities.[10] They speculated that such students experienced a damaging sense of frustration and failure during assessment administration, which teachers felt accounted for the heightened incidence of outbursts and disturbances these students exhibited immediately after the assessments.

Interviews with special education students right after the MSPAP administration suggest that the experience was upsetting to many. One student said he assumed ". . . they're testing me to see how bright I am" and that he found the MSPAP almost impossible because the test had been "worded too difficult." Assessment accommodations, such as assessment scheduling, setting, and special equipment, for students with disabilities are tied directly to their IEPs. However, wording of the assessments is not changed for any student population; and, since some special education students are separated from their peers during MSPAP administration, they do not receive the benefit of working with stronger students during the group portions of the test.

---

[9] MSPAP tests are administered for 1 hour and 45 minutes over 5 days in grades 3, 5, and 8. CRES tests take anywhere from one to five periods to administer in each discipline at each grade level, and Maryland functional tests are administered one day each year at each grade level.

[10] All special education students, except those with "profound deficiencies" must take the CRES and the MSPAP; and English-as-a-second-language students are exempt from the assessments for only one year.

At the other end of the learning spectrum, some teachers were concerned that honors students felt too much "time pressure" because of the lengthy responses needed to answer open-ended questions.

Impact on Parents

Walters parents felt they had received enough information about both the CRES and MSPAP assessments to inform them of assessment content and goals. The WCPS officials have tried to keep the parents and the community informed by publishing brochures on the educational reforms they have undertaken. WCPS parents thus seem strongly supportive of the CRES rubric scoring, which they said provided clearer information about the academic standards being set for their children and about their children's strengths and weaknesses with regard to those standards than did traditional letter grades. They complained, however, that, unlike CRES scores, individual MSPAP scores were not currently available in a home-report format.

CRES and MSPAP Future Plans

WCPS officials feel more confident about the future of CRES than about the accountability-oriented MSPAP. They feel public educational campaigns are needed to explain the high standards and the goal-oriented nature of the MSPAP, especially since students' poor showing over the last two years has generated something of a "public relations shock." Even in Walters County, where students averages were better than the overall state average, media response to scores was negative, ". . . with headlines like 'Local Students Flunk State Test'," said the board member.

District officials believe that some political groups add fuel to the "public fire" (in some cases without adequate or factual knowledge of the system) by suggesting that the MSPAP (and CRES) "sacrifice" basic skills in the interest of "liberal" values. The district officials believe that the best way to counter such criticism is to define learner outcomes in terms of content proficiencies. It should also be noted that the debate still rages about whether it is wise for the state to legislate blanket remedies to perceived local educational problems.

District officials, despite some negative reaction from various parties, are pushing ahead with more progressive changes to CRES. They are piloting a cross disciplinary high school portfolio that is to contain five pieces of best work; and Physical Education and the Arts are being added to the CRES summative assessments. To improve staff development and staff support, and to lessen the impact of test time on instruction, officials plan to add 5 staff days to the school year by increasing the length of the year from 185-to-190 days.

Epilogue

Aggregate data for the 1993-94 MSPAP administration showed that satisfactory standards were not met for any grade level or content area.[11] MSDE intends to continue to work with the local school systems to improve instructional quality to achieve the set standards. In the meantime MSPAP received favorable publicity when *The Sun* asked some parents and other citizens (including a Baltimore city councilman) to take the MSPAP. MSDE administered the fifth grade MSPAP assessments to these adult citizens, who found the assessments challenging, but not inappropriate.[12] In addition, various schools planned different types of events, such as bringing in a helicopter, to motivate their students for the 1994-95 assessments.[13] The results of this assessment administration will not be available for several months.

Conclusions

This case study examined the implementation of both district and state performance assessment systems in a Walters County, Maryland middle school. The purpose of the Maryland School Performance Assessment Program is to raise standards of academic achievement in all Maryland schools and to make individual schools accountable for meeting those standards.

While Walters County school district officials agree with the principles of the MSPAP, they are concerned about certain specific features of the assessment. They said that although the MSPAP represents an enormous investment in administrative and lost instructional time, it cannot be used for purposes of instructional feedback, since results are reported a full year after assessment administration. However, MSPAP is useful in that it can be used to validate the district's own assessment system, CRES. (MSDE plans to report the MSPAP results earlier than it has in the past. In addition, it plans to better inform teachers about the Maryland learning outcomes.)

The district's CRES, on the other hand, has been designed specifically for the purpose of instructional feedback and to ensure that students learn the district's new "Essential Curriculum." For this reason, the CRES involves the administration of "formative," as well as "summative," examinations that teachers can use to monitor student progress and to adjust instructional techniques. Whereas the MSPAP is administered only in four traditional academic subjects at only three grade levels, the CRES is administered at every grade level and in every academic discipline.

The impact of these assessments has been both positive and negative. Teachers believe the CRES has improved their teaching, but they complain that it is difficult to reliably score the assessment. District officials agree that the CRES is psychometrically fragile, although recent checks of inter-rater reliability have been quite positive. As a result of their experiences, district officials consider the test a "work in progress," not to be used in the foreseeable future for accountability purposes. The impact of MSPAP in the classroom is much less obvious.

---

[11] *Maryland School Performance Report, 1994, State and school systems.* Maryland State Department of Education, Baltimore, MD.

[12] *Matching wits with Md 5th graders* (May 11, 1995). The Sun.

[13] *Rallies, hoopla kick off key Md. exam.* (May 5, 1995). The Sun.

Virtually every constituency interviewed at Walters (teachers, administrators, students, and parents) believes that both the MSPAP and the CRES examinations are more interesting, relevant, and challenging than are traditional minimum competency tests administered throughout the state. Some teachers indicate that the tests improved the real-world application of their instruction and forced them to focus on teaching reasoning skills and interdisciplinary writing. Despite concern that both assessments not waste instructional time, the overall view of these new assessments is largely a positive one; Walters County educational community feels that they represent not only a departure from old ideas and policies that are no longer in the best interests of students, but also offer a view into the future of education.

APPENDIX A

MSPAP Grade 8 Mathematics Task

MSPAP
PUBLIC RELEASE TASK

Birth Dates

Grade 8

Mathematics

NOTICE

Use of this and all non-secure public release tasks to train educators and to familiarize students with the Maryland School Performance Assessment Program is encouraged. However, copyrighted text material cited on the acknowledgments page is protected by federal copyright law. Before duplication and dissemination of copyrighted material on an extensive basis, consideration should be given to securing appropriate copyright permission.

Maryland State Department of Education
July 1994

ACKNOWLEDGMENTS

The Maryland State Department of Education would like to thank:

- The hundreds of local school system staff who are members of the Content Coordinating Teams, the Test Administration Team, the Scoring Advisory Team, and the Task Development Teams that developed these and all MSPAP tasks.
- Its specialists from the Division of Instruction, the Division of Planning, Results, and Information Management, and other divisions for their assistance in the creation of MSPAP.

STUDENT RESPONSE BOOK

Task 2 Day 1
Topic: Birth Dates

Some people believe that the time of year you are born influences some of your preferences. Perhaps if you knew who shared the same birth month, you could see if you have similar tastes.

PERCENT OF PEOPLE BORN IN EACH MONTH OF THE YEAR

January	5	May	6	September	15
February	6	June	7	October	12
March	4	July	9	November	9
April	6	August	10	December	11

1 Choose an appropriate graph for the birthday data. Construct the graph below.

2

STEP A In what month were you born?

____

STEP B If there are 538 students in your school, about how many are likely to share your birthday month?

____

GO ON

4

STEP A

Imagine you found 28 people in your grade born in the same month as you. Would you expect to find that two or more of these people share the same birthday? Circle one: YES NO

Explain your prediction.

STEP B

You have decided to check the prediction you made above. Explain the steps you would take to collect data and how the data could be displayed.

STEP C

Justify your display choice.

EXAMINER'S MANUAL

400

When students have completed activity 1, Step D, they are ready to again report their data. Record these data on the classroom chart, Class Data on the Neutralization of Waste Water, according to the chart, Group Reporting Schedule for Class Data on Neutralization of Waste Water.

Group Reporting Schedule for
Class Data on Neutralization of Waste Water

TOTAL VOLUME OF NEUTRALIZER	REPORT #1	REPORT #2	REPORT #3	MEAN
5 mL	G1 S1	G4 S1	G3 S1	
10 mL	G5 S1	G2 S1	G6 S1	
15 mL	G1 S2	G4 S2	G5 S2	
20 mL	G6 S2	G2 S2	G3 S2	
25 mL	G1 S3	G5 S3	G7 S1	
30 mL	G6 S3	G7 S2	G3 S3	
35 mL	G4 S3	G2 S3	G7 S3	

G = GROUP
S = STEP

SAY **Copy the data from the classroom chart, Class Data on the Neutralization of Waste Water, into your chart on page 3 of your Student Response Book.**

Point to the chart. Allow time for students to copy these data into their chart, Class Data on the Neutralization of Waste Water, in their Student Response Books.

SAY **You will now complete the task working on your own.**

Task 2, Day 1 Topic: Birth Dates

SAY **Now you are going to work on a task dealing with birth date preferences. Open your Student Response Book to page 11 and find task 2. You will have 30 minutes to complete this task. You may now begin.**

MSPAP
PUBLIC RELEASE TASK

Birth Dates

SCORING GUIDE

INTRODUCTION, SCORING TOOLS,
AND SAMPLE RESPONSES

Grade 8

Mathematics

Maryland State Department of Education
July 1994

ACKNOWLEDGMENTS

The Maryland State Department of Education would like to thank:

- The hundreds of local school system staff who are members of the Content Coordinating Teams, the Test Administration Team, the Scoring Advisory Team, and the Task Development Teams that developed these and all MSPAP tasks.
- Its specialists from the Division of Instruction, the Division of Planning, Results, and Information Management, and other divisions for their assistance in the creation of MSPAP.

INTRODUCTION TO SCORING GUIDE

The following collection of scoring tools comes from an operational scoring guide used by the team of Maryland teachers who scored this task from the 1992 MSPAP. It is important to remember that such scoring guides are only a supporting resource used to train teachers to score MSPAP. The guides, along with several training sets, are reviewed over two days of training. During this time, the team coordinator and team leader identify essential criteria for each score decision, and carefully explain why the particular sample responses were selected to represent a given score point. Many of the samples in these training materials are not immediately obvious examples of a given score point. They have been selected to permit discussion of a wide range of scoring issues and factors that contribute to a score decision. During training, teachers have an opportunity to compare many sample papers and seek guidance on the score decision process. Only after this training do they score several qualifying sets, packets with pre-established "true" scores against which their performance is compared. While the primary purpose of these qualifying sets is to ensure that teachers can score with sufficiently high exact agreement before beginning the actual scoring process, these responses also provide further opportunities to explore and clarify factors contributing to score decisions.

It is likely, therefore, that some of the scores for the sample responses that follow will not be immediately clear to readers. They all have been selected, however, based on consensus by a team of Maryland educators, backed up by the scoring contractor's senior staff. It will be most helpful if you use them to get a general "feel" for what characterizes both satisfactory and excellent performance. Remember that a score of "1" always means "at least satisfactory." While there may be some ideas or information of merit in some sample "0's," in the judgment of Maryland teachers those responses were not yet satisfactory. They might be too vague, contain too much erroneous information, or require the scorer to make too many inferences about intended meaning. While scoring tools vary in terms of the number of score points that may be assigned, the highest score point is reserved for excellent responses. These are responses that fully address the demands of the particular activity and reflect the conventions of the discipline being assessed.

Sometimes, when scoring performance assessment tasks, readers use a single scoring tool to assign a score for multiple outcomes, either from the same content area or from different content areas. This is called "simultaneous scoring." In the case of this task, some of the scoring tools capture both mathematics process and content outcomes. In other tasks, in instances when student performance in one outcome area may differ from performance in another area assessed by the same activity, two different scoring tools are used. This is called "sequential scoring." Sometimes mathematics content and process scores are derived from sequential scoring.

The following is a summary of the assessment activities (i.e., items) in this task and the areas of knowledge and skills assessed by each activity. Scoring tools and actual student responses follow.

Activity	Outcome Assessed
1	Statistics (M outcome 10)
2A	DO NOT SCORE*

Task 2/Day 1
SCORING TOOLS FOR 8BM10
Topic: Birth Dates
Page 11

Activity 1

This response indicates the ability to organize data obtained from a display.

2 = Complete and correct graph. Must contain all features listed below

1 = A and B are correct, in C not more than one section of the graph is incorrectly proportioned and in D data from not more than one month is inaccurately recorded.

OR

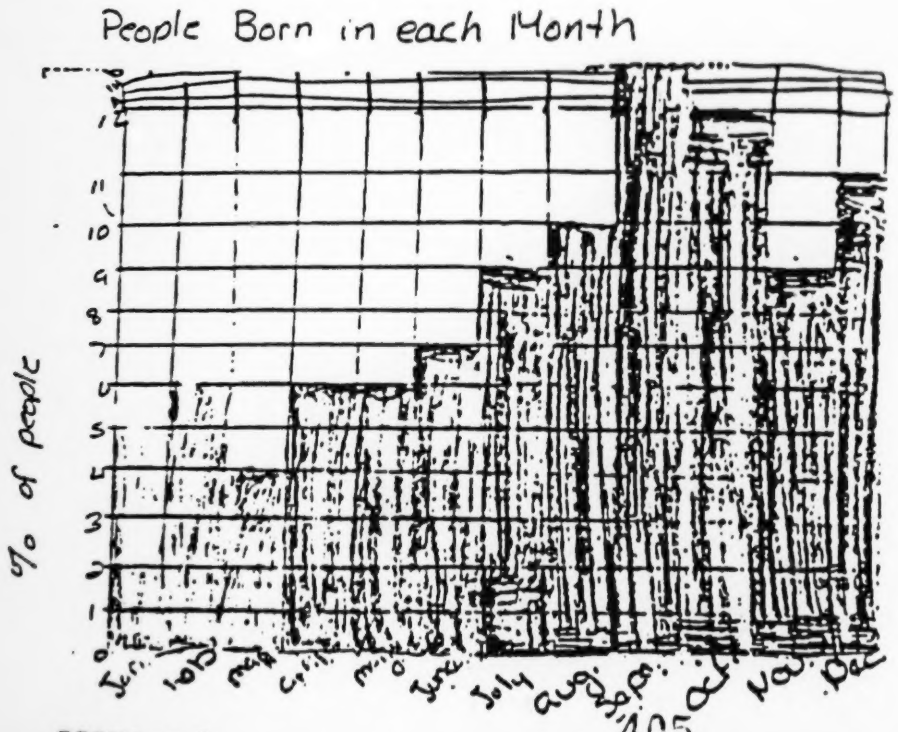
C and D are correct and either A or B are missing

0 = other

Answer cue:

A. correct axis labels: month and percent of people
B. title appropriate to graph
C. equal and/or proportional intervals marked on axes
D. correctly plotted data

Student Responses



score = ?

Activity 2C (Depends on 2A)

This response indicates the ability to communicate the mathematical procedure used in arriving at an answer.

1 = correct

0 = other

Answer cue:

Answer consistent with month of student's birth.
Multiply percent of birthday month times 538. Rounded to nearest whole number
OR
If student's birth month is: then response must reflect the following arithmetic:

January	.05 x 538; 5/100 x 538; 1/20 x 538
February	.06 x 538;6/100 x 538; 3/50 x 538
March	.04 x 538; 4/100 x 538; 1/25 x 538
April	.06 x 538;6/100 x 538; 3/50 x 538
May	.06 x 538;6/100 x 538; 3/50 x 538
June	.07 x 538; 7/100 x 538
July	.09 x 538; 9/100 x 538
August	.1 x 538; 10/100 x 538; 1/10 x 538
September	.15 x 538; 15/100 x 538; 3/20 x 538
October	.12 x 538; 12/100 x 538; 3/25 x 538
November	.09 x 538; 9/100 x 538
December	.11 x 538; 11/100 x 538

Activity 4A

This response indicates the ability to use probability in making predictions and to communicate the mathematical reasoning used in solving a problem with an open-ended answer.

2 = complete explanation including all possible outcomes

1 = partial explanation indicating high probability

0 = other

Answer cue:

Yes. Answer should include the idea that while it is possible that each of the 28 people are born on different days; it is probable that 2 or more people do share the same birthday. It should also include some support for this idea.

Exemplary Response:

My month has 30 days. If 20 students were born on different days, then the chance would be 10/30 that the 21st would not share a birthday, 9/30 for the 22nd, 8/30 for the 23rd, etc. Except for the 1st person, everyone's chance is less than 1. 10/30 x 9/30 x 8/30 = 2/75. So the chance that everyone has a different birthday is much less than 2/75.

Student Responses

Yes

I have two friends who have the same birthday.

Feasible but not scorable

Activity 4C (Depends on 4B)

The response indicates the ability to communicate the reasoning used in choosing an appropriate data display model.

1 = The response must be consistent with choice of display.

0 = other

Answer cue:

- frequency table: a quick method to display data as the data is likely to be spread out
- line plot or bar graph: shows clusters and gaps at a glance

Student Responses

A graph is a easy way to look at and compare results to a problem.

score = 1

APPENDIX B

MSPAP Results

STATE OF MARYLAND
DISAGGREGATED DATA, DATA-BASED AREAS

School Year 1993-1994

The following data have been disaggregated for the data-based areas included on pages 10 and 11 of this report. Please refer to the introductory section of this report for definitions of the various data-based areas. (Data in four Baltimore City high schools--Frederick Douglass, Patterson, Walbrook, and Northwestern--cannot be verified; therefore, data from those schools have been excluded from all areas of Functional Tests--Grades 9 and 11; Attendance Rate, Grades 7-12; and Dropout Rate.)

STUDENT PERFORMANCE
ASSESSED STUDENT KNOWLEDGE

MARYLAND FUNCTIONAL TESTS PERCENT PASSING	STANDARD %		American Indian/ Alaskan Native		Asian/Pacific Islander		African American		White (not of Hispanic origin)		Hispanic		All Races	
	Excellent	Satisfactory	M	F	M	F	M	F	M	F	M	F	M	F
GRADE 9 STATUS														
Reading	97	95	98.0	94.5	97.8	98.3	93.0	96.7	98.3	99.2	94.9	97.3	96.4	98.3
Mathematics	90	80	81.6	80.7	93.6	95.4	62.5	67.0	89.4	90.7	79.3	80.5	80.3	82.5
Writing	96	90	89.6	92.5	94.2	97.6	77.9	86.8	90.1	95.5	87.2	95.1	86.2	92.7
Citizenship	92	85	87.0	86.8	92.8	93.0	75.1	78.2	91.0	90.6	77.0	80.5	86.3	87.0
GRADE 11 STATUS														
Reading	99	97	97.4	100.0	99.1	99.6	99.3	99.7	99.8	99.9	98.9	99.7	99.6	99.8
Mathematics	99	97	95.0	95.2	98.4	98.6	90.5	92.5	98.6	98.7	92.2	94.6	96.0	96.7
Writing	99	97	89.7	100.0	98.4	98.5	96.2	98.5	99.1	99.6	96.2	96.3	98.1	99.2
Citizenship	99	97	85.0	97.6	96.4	95.3	93.7	93.7	98.3	97.7	90.2	90.2	96.6	96.2
Passed all tests	96	90	77.5	95.2	94.5	93.9	85.2	87.8	96.8	96.7	84.2	85.8	92.9	93.5

MARYLAND SCHOOL PERFORMANCE ASSESSMENT PROGRAM (MSPAP)		STANDARD %		American Indian/ Alaskan Native				Asian/Pacific Islander				African American				White (not of Hispanic origin)				Hispanic				All Races			
				M		F		M		F		M		F		M		F		M		F		M		F	
				% at		% at		% at		% at		% at		% at		% at		% at		% at		% at		% at		% at	
		Excellent	Satisfactory	EX	SAT	EX	SAT	EX	SAT	EX	SAT	EX	SAT	EX	SAT	EX	SAT	EX	SAT	EX	SAT	EX	SAT	EX	SAT	EX	SAT
GRADE 3	Reading	25	70	3.0	26.0	5.5	35.2	4.3	38.2	7.3	47.0	0.5	10.6	1.2	17.4	3.6	34.3	6.3	43.2	1.2	22.7	1.9	25.9	2.6	26.6	4.5	34.5
	Mathematics	25	70	1.0	22.5	2.2	25.8	6.2	48.0	6.3	51.9	0.4	12.5	0.5	15.9	4.0	41.6	4.0	45.5	0.8	24.4	1.6	23.9	2.8	32.2	2.9	35.5
	Social Studies	25	70	1.0	24.5	1.1	30.1	1.2	39.1	2.1	47.3	0.1	11.6	0.3	17.7	1.3	37.6	2.2	45.4	0.5	21.5	0.8	22.9	0.9	29.0	1.5	35.8
	Science	25	70	1.0	27.5	2.2	30.1	5.7	45.3	8.5	52.4	0.6	13.4	1.2	18.0	5.2	42.2	7.0	46.4	2.4	24.2	1.7	26.1	3.7	32.7	5.0	36.8
	Writing	25	70	7.8	30.4	11.8	40.9	10.7	42.6	15.0	53.7	2.1	16.7	4.3	25.4	7.6	36.5	13.5	47.9	3.0	25.7	6.0	29.1	5.9	30.2	10.3	40.3
	Language Usage	25	70	6.9	27.5	10.8	35.5	16.8	48.0	28.1	63.0	2.6	13.8	5.6	24.0	10.7	34.3	19.0	48.5	5.0	20.5	8.0	26.6	8.2	27.9	14.7	40.4
GRADE 5	Reading	25	70	0.0	18.9	8.3	33.3	4.7	37.0	9.6	54.5	0.6	10.1	1.5	19.3	3.0	30.8	6.4	45.1	1.1	17.7	1.7	25.1	2.2	24.1	4.7	36.2
	Mathematics	25	70	5.6	37.4	10.2	39.8	18.2	60.7	17.6	67.5	1.1	15.8	1.4	20.5	11.1	52.4	11.2	55.5	4.4	31.8	3.1	31.4	8.0	40.5	7.9	43.5
	Social Studies	25	70	0.0	22.4	5.7	36.4	4.7	39.4	8.4	49.8	0.8	11.3	1.1	16.0	5.3	39.5	7.6	45.6	1.8	23.7	2.4	24.0	3.8	30.1	5.3	35.3
	Science	25	70	1.9	32.7	5.7	37.5	8.9	53.6	11.8	61.0	0.6	12.9	1.0	18.4	7.0	47.7	9.5	53.0	2.8	27.4	3.5	28.9	4.9	36.4	6.6	41.0
	Writing	25	70	7.5	29.9	17.0	33.0	14.3	38.3	24.4	53.7	2.4	13.8	4.2	20.7	11.6	36.2	17.1	46.4	6.5	24.1	6.6	24.8	8.7	28.9	12.8	37.4
	Language Usage	25	70	10.4	23.6	23.9	43.2	21.4	45.1	34.5	63.9	3.6	13.2	7.1	22.7	15.5	36.6	26.1	50.5	8.3	23.6	10.3	27.7	11.7	29.1	19.6	41.0
GRADE 8	Reading	25	70	0.0	23.2	0.0	27.1	2.9	32.4	4.9	49.1	0.2	6.4	0.7	13.8	1.8	24.5	3.4	37.4	0.8	11.9	0.9	18.1	1.3	18.6	2.4	29.3
	Mathematics	25	70	4.8	32.1	3.5	32.6	14.0	63.2	13.9	62.9	0.4	13.1	0.6	17.4	7.3	51.8	7.6	54.4	1.4	24.8	2.0	24.7	5.1	39.0	5.3	41.4
	Social Studies	25	70	2.4	26.2	1.2	25.6	5.0	47.0	8.4	53.2	0.3	9.6	0.9	17.3	3.7	36.4	5.7	46.4	1.0	16.6	1.0	20.3	2.6	27.6	4.1	36.0
	Science	25	70	6.0	34.5	7.0	33.7	17.1	59.8	19.6	62.0	0.9	13.1	2.0	19.6	12.1	49.1	14.0	54.2	4.9	25.8	4.4	25.8	8.5	37.3	9.9	42.0
	Writing	25	70	10.7	38.1	10.5	38.4	16.9	58.2	24.8	69.4	1.9	18.4	5.8	34.1	9.8	44.5	19.5	61.3	3.4	26.8	7.4	36.7	7.4	36.2	14.7	51.7
	Language Usage	25	70	9.5	35.7	12.8	40.7	21.9	59.4	35.1	71.0	2.1	16.4	7.0	33.8	11.8	43.7	23.5	62.2	4.5	23.5	9.2	34.3	8.8	35.0	18.0	52.1

* Fewer Than 20 Students

BEST COPY AVAILABLE

APPENDIX C

CRES Scoring Rubric

ADMINISTER THE AGGREGATE ASSESSMENT

- o Teach the social studies essential curriculum
- o Use texts, maps and globes, multimedia, visitors and field experiences
- o Employ cooperative learning, thinking strategies and graphic organizers
- o Integrate social studies with other disciplines where possible
- o Administer assessments when appropriate
- o Read performance directions thoroughly before administration
- o Follow directions for teacher and students
- o Use attached materials, organizers and rubrics
- o Only score assessment
- o Score with attached rubrics
- o Give rubric scores of 3, 4 for proficient; and 1, 2 for nonproficient
- o Record student scores on suggested classroom record chart
- o Reteach if students receive a nonproficient score during class or tutoring time
- o Administer the reteaching assessment. Score with attached rubric. Record retest score.
- o Give Latin American Objective Test (50 items; no reteaching)
- o File written responses
- o Determine social studies mastery at year end by using cumulative score
- o Destroy papers end of year except for those requested for interscorer reliability
- o ADJUST ACTIVITIES TO MEET SPECIAL INDIVIDUAL NEEDS

Social Studies Assessment: PAMPHLET/BOOKLET 2-3 Class Sessions

Grade Level: 8

Unit of Study: River Valley and Mediterranean Civilizations

Essential Curriculum Objectives:

- o Analyze historic cultures by location, politics, society, economics, religion, intellectual achievements and military organization
- o Interact positively through self-respect and enhanced appreciation of social diversity
- o Work cooperatively in groups and accept group decisions
- o Interpret, evaluate, organize and use information by observing, investigating, listening and reading
- o Gather information, think critically and solve problems
- o Relate historical problems to modern events

Teacher Notes:

1. This assessment may be given at any time during the first 3 terms.
   The assessment may be given during or after the regular unit of study on River Valley OR Mediterranean Civilizations.

2. A graphic organizer is available:
   - -- one civilization will be assigned
   - -- may be assigned any time before, during or after unit of study
   - -- may be an ongoing assignment (process) during unit or as a culminating activity
   - -- may be collected and stored by teacher until assessment

3. Teacher may choose which civilization(s) will be assessed or may allow students to choose after the study of several civilizations.
4. Teacher may display diffeent kinds of pamphlets as examples.

*Assessment: As an historian, you will create a pamphlet/booklet for an 8th grade class on one ancient civilization which includes:

1. a map showing the location of that civilization
2. descriptions and illustrations of at least three contributions from different aspects of an ancient civilization
3. explanations of how each contribution has influenced (affected) later civilizations

Pamphlet/Booklet (Student Choice) Specifications

Should include:

o Cover sheet with civilization's title and an "interesting" illustration
o Map depicting major geographic/cultural features for a time period
o Three contributions (one per page/panel) from three different topics (total = 3); for example, goverment/society/religion

Requirements:

o Illustration(s) of contribution (pictures from magazines or student drawings)
o Written in paragraph form including:
  . A description of the contribution
  . An explanation of what the contribution tells about the ancient civilization
  . An explanation of how the contribution has influenced (affected) later civilizations

*Problem Solving Rubric

*Reteaching Assessment: Students will write information in paragraph form or tell their teacher about their civilization. They will use the graphic organizer for sharing at least four contributions of this civilization which they think are interesting or important.

ADJUST ASSESSMENT ACTIVITIES TO MEET SPECIAL NEEDS.

PROBLEM SOLVING RUBRIC

Proficient

4 = The student demonstrates an exceptional understanding of the problem and uses effective problem solving strategies.

- Data collected, well organized and properly displayed
- Any inaccuracies are minor and do not affect successful solution
- Response is supported with many, varied (and unusual) details

3 = The student demonstrates an understanding of the problem and uses appropriate problem solving strategies.

- Data is collected, organized and displayed
- Any inaccuracies/omissions are minor and have minimal effect on the solution
- Response is supported with some detail

Nonproficient

2 = The student demonstrates a partial understanding of the problem and attempts to use a problem-solving strategy.

- Some data is collected but may be ineffectively organized or displayed
- Inaccuracies and/or errors interfere with solution
- Response is not supported by information and may be too general or incorrect

1 = The student is unable to demonstrate an understanding of the problem, but attempts a response.

- Response may be confusing, irrelevant and/or incorrect

MANZANITA

The Arizona Student Assessment Program: Manzanita High School April 27-29, 1994

THE ARIZONA STUDENT ASSESSMENT PROGRAM: MANZANITA HIGH SCHOOL

Introduction

Signed into law by Governor Rose Mofford in 1990, the Arizona Student Assessment Program is intended to guide education reform in Arizona. The legislation, which was generated through the collaborative efforts of Arizona's Joint Legislative Committee on Goals for Educational Excellence, the State Board of Education, and the Arizona Department of Education (ADE), establishes goals for the public elementary and secondary educational system and sets out methods for assessing achievement of those goals. Specifically, the legislation calls for the development of:

- curriculum standards,
- a high-quality, broad-based assessment system,
- a comprehensive reporting system,
- a graduation rate tracking system, and
- a postsecondary tracking system.

The primary focus of this case study is the development and implementation of the Arizona Student Assessment Program, or "ASAP" (the second element of the legislation), and its impact at one high school. ASAP is designed to measure student progress toward attaining the state's designated *Essential Skills*, the curriculum standards developed by the ADE and approved by the Joint Committee. Administered statewide for the first time during the 1992-93 school year, ASAP assesses all Arizona students in reading, writing, and mathematics at grades 3, 8, and 12. The ADE has plans to expand ASAP to include other subject areas (i.e., science and social studies) and to make satisfactory performance on the ASAP a requirement for high school graduation. ASAP, then, has two overall objectives: to *certify* student achievement and to *establish* standards of student achievement to which communities can hold their schools accountable.

The top performing school district in 1993 and 1994 on the 12th grade version of the ASAP was the Desert View Union High School District, located in suburban Phoenix and Phoenix proper. Manzanita High School, the focus of this case study, is one of nine high schools that comprise the district — and one that reflects both urban and suburban cultures.

During the 1993-94 school year, Manzanita High School served approximately 1,036 ninth through twelfth graders. These students were primarily white (78 percent); other students were Hispanic (15 percent), African American (4 percent), Asian American (2 percent), and Native American (1 percent). Students are from middle to lower-middle class families, and school staff report that the student population has become increasingly transient over the past five years or so; they also report that minority representation in the student body has grown over the same period.

Located in the center of the Desert View district, Manzanita's enrollment area is pulled on by schools to its west, north, and east, and the construction of a large shopping mall across the street from the school in the late 1970s further constricted the school's potential attendance zone. Consequently, the school that served over 2,500 students 20 years ago experienced its lowest enrollment ever in 1993-94.

About half of Manzanita graduates go on to college, with one-third of these going to four-year colleges or universities and two-thirds to community colleges. The school's annual dropout rate has ranged from 6 to 13 percent between 1987 and 1994. According to both students and teachers, some students at the school belong to street gangs. However, teachers say that, on the whole, the gangs do not present "substantial problems," and violent incidents are rare.

The Manzanita teaching staff of 56 is experienced, and many teachers have been at the school for more than 20 years.

Participants

More than a dozen individuals associated with Manzanita High School participated in this case study. In addition, two district administrators and one official at the Arizona Department of Education answered questions by telephone about ASAP and the Desert View district. The roles of individuals interviewed for this case study are identified in Exhibit I.

EXHIBIT I

Study Participants

- Deputy Associate Superintendent, Arizona Department of Education
- Associate Superintendent for Curriculum and Instruction, Desert View Union School District
- Director of Research, Desert View Union School District
- School Board President
- Principal
- Assistant Principal
- Six Teachers
  - English, Social Studies, Mathematics, Foreign Languages; some of the teachers teach seniors; all of the teachers have been involved in the administration of ASAP; one teacher is the Manzanita High School representative to the Arizona Education Association
- Five Students
  - Seniors of varying achievement levels and future plans, all of whom took the 12th grade administration of the ASAP in January 1994
- Two Parents of students who have taken the ASAP

Observations

In addition to the interviews, lessons in several classrooms were observed. Observation of the administration of ASAP was not possible.

State Context

The Legislative Mandate and the Purposes of ASAP[1]

During the second half of the 1980s, the Arizona legislature, like many of its counterparts in other states responding to the publication of *A Nation at Risk*, put its concern about the quality of public education in the state on the "front burner." Legislation was enacted in 1987 to establish the Joint Legislative Committee on Goals for Educational Excellence, whose mandate it was to identify goals to guide Arizona educators into the future. Working in partnership with the Arizona Board of Education and the Arizona Department of Education, the Joint Committee found that the Department of Education had been developing curriculum frameworks, called *Essential Skills*, in academic subject areas (language arts, mathematics, science, and the social sciences). The Committee further found that the goals for student outcomes contained in the *Essential Skills* documents adequately articulated the vision the Committee had in mind for Arizona's students. Considering this background, the Committee recommended to the Legislature that the *Essential Skills* be adopted in lieu of developing new frameworks and goals.

However, when the Committee turned its attention to the state-administered assessments of student achievement, it found that neither of the two instruments in use — the Iowa Test of Basic Skills (ITBS) and Arizona's Continuous Uniform Education System assessments (CUES) — corresponded well to the *Essential Skills*. In fact, the Joint Committee's evaluation of the match between ITBS items and *Essential Skills* found that only 26 percent of the *Essential Skills* were tested on the assessment; for the CUES, administered differentially by districts, the state had no data.

The Committee recommended that new assessments specifically aligned with the *Essential Skills* be adopted. The Arizona Student Assessment Program was subsequently established by act of the State Legislature and signed into law by Governor Mofford in May of 1990. The purpose of ASAP, which encompasses both state-administered performance assessments and, at the district level, District Assessment Plans and Essential Skills Completion Reports (described below), is to measure student progress toward attaining the state's *Essential Skills*.

Development of ASAP

The Arizona Department of Education contracted with the Riverside Publishing Company to develop the ASAP assessment instruments. (Riverside also is the publisher of the two assessments used by the state or many of its districts, the ITBS and the Test of Achievement and Proficiency (TAP).)

---

[1] This section is drawn primarily from Easton, L.B., & Koehler, P. (1994). Arizona's educational reform: Creating and capitalizing on the conditions for policy development and implementation. In Pelavin Associates, Inc. (ed.) *Implementing Performance Assessment: Promises, Problems, and Challenges*. Washington, D.C.: Author.

ASAP's Forms A, B, C, and D

For the ASAP, Riverside was charged with the task of developing four forms of assessment instruments. Form D, the form to be used for the state's formal testing program, would be administered to students in grades three, eight, and twelve.

The three other forms — Forms A, B, and C — were to be developed for optional district and school use. Form A, also intended for use with third, eighth, and twelfth graders, was developed first; it is used by districts, schools, and teachers as they desire (e.g., districtwide assessments, instructional tools, and practice tests). Forms B and C also are intended for use in the intervening grades at the discretion of districts and schools.

Revision and Pilot Testing of ASAP

The Arizona Department of Education distributed Form A to all 220 Arizona school districts during the 1990-91 school year. The Department of Education specifically solicited teachers' feedback to the assessment tasks on Form A. The feedback received by ADE and Riverside led to revisions in Form A and guided the development of Form D (again, the form to be used for the statewide administration of ASAP). As an example of feedback, teachers commented that the math assessments "did too much thinking for students," and so they were revised.

The ASAP Form D assessments were then piloted during the 1991-92 school year. All Arizona school districts and all non-exempt students participated in the pilot. An evaluation of the validity and reliability of ASAP was conducted by Riverside, but that report was not available for this study. According to the ADE Deputy Associate Superintendent, interrater reliability for the scoring process (each assessment was scored by two scorers) was high.

Full-Scale Implementation

The first full-scale administration of the ASAP took place in March 1993. In all, just under 129,000 third, eighth, and twelfth graders took the assessment.[2] Just over 5,000 of those students took the assessment under mediated conditions (e.g., adjusted scheduling, translation, reading to the student), primarily because the students were served in programs for children requiring special education services or programs for children with limited proficiency in English.

Twenty-two school districts across the state volunteered to serve as scoring sites for the 1993 ASAP. In all, 675 Arizona teachers served as scorers at the sites. Those teachers, together, scored about 25 percent of the completed assessments; the remaining 75 percent of the assessments were scored by Measurement, Inc., located in North Carolina. Again, each assessment was scored by two scorers; and all scorers were trained by Riverside staff.

---

[2] This number actually represents the number of assessments *scored* following the March 1993 administration. The number of unscored assessments is not available.

Characteristics of ASAP

As previously mentioned, the Arizona Student Assessment Program tests all third, eighth, and twelfth grade students who are not exempt (by virtue of provisions in their IEPs) from the program in the areas of reading, writing, and mathematics. The format is a performance event; students construct responses to prompts contained within assessment booklets (one booklet for each of the three assessments). Teachers are responsible for administering the assessment, and the district is responsible for collecting the assessment booklets and sending them back to the ADE to be scored.

Performance Events

The assessment at each grade level maintains a single theme across the three sections; for instance, in 1994, the twelfth grade assessment followed a "consumer decisions" theme, while in 1993, the twelfth grade theme focused on rain forests. Bridging activities designed to focus students on the assessment activities and subject matter occur between sections of the assessment. Students are allowed two hours to complete each section of the assessment — for a total of six hours for the entire assessment (though, as is explained below, there apparently has been some confusion during the first two administrations of the exam as to how rigidly the time limits should be enforced).

Tasks on all three parts of the assessment — reading, writing, and mathematics — require students to construct responses to prompts. For example, a writing assessment might ask students to write a rough draft of a brief essay and then a final draft on a topic related to the assessment's theme. One example of an eighth grade ASAP mathematics task might ask students to design a playground according to particular space and cost specifications; students might be asked to justify their responses in writing. A reading assessment for eighth graders would typically require students to read a story or essay and answer a dozen or so questions about the story with brief (i.e., one to three sentences) responses. One sample eighth-grade mathematics task appears in Appendix A.

Students who require mediated administrations of the assessment receive them, according to their Individualized Education Programs. Types of mediation offered during the 1993 administration of the exam included adjusted scheduling and setting of administration, translation, mediator reading to the student or taking dictation from the student, and revision of directions. Many students who took the exam under mediated conditions received more than one type of mediation.

Assessments are aligned to Arizona's *Essential Skills* areas. The *Essential Skills* for twelfth-grade reading, writing, and math appear in Appendix B. Each year, the assessment "audits" students on a subset of the Essential Skills. As a result, the state does not purport that results on the assessments are comparable from year to year. For example, in the 1993 administration of the twelfth-grade ASAP, the following *Essential Skills* were assessed:

- **Reading** — (1) identify the purpose and audience of the communication; (2) identify response wanted from audience; (3) evaluate methods used by sender of communication; (4) identify author's strategy; (5) relate to own experience; and (6) identify the author's point of view.

- **Writing** — The student must write a letter that (1) makes clear the audience and purpose for which the communication was intended; (2) meets the needs of the audience for that purpose in an efficient and effective way; (3) contains no gaps, omissions, or assumptions that could impede comprehension of the communication; (4) follows the standard style for a business letter; (5) shows evidence of editing and proofreading final drafts so that errors in spelling, punctuation, capitalization, and usage do not impede comprehension; (6) has proof that is "tied-in" to each point and to the thesis, that is, the writer explains how the proof truly proves the point; and (7) is organized effectively and includes an introduction that leads into the argument and rousing conclusion that leaves no doubt about the issue in the reader's mind.

- **Mathematics** — (1) use deductive reasoning to generate conclusion; (2) distinguish between inductive and deductive reasoning and explain how each is appropriately used; (3) use the four arithmetic operations on rational numbers (expressed as common fraction, decimal fractions, or percents) with accuracy and reasonable speed, using pencil and paper or calculator; (4) solve problems using ratio and proportion, percent, or direct and inverse variation; (5) explain variability of statistical measures in terms of sampling process or population differences; (6) select and use appropriate principles of counting collections and arrangements of objects or outcomes of sequential procedures; (7) use statistical reasoning to test hypotheses informally; (8) write the linear relationship between two variables by inspection of its graph or of the set of its ordered pairs; and (9) recognize when conditions of definitions are satisfied.

Scoring Rubrics

Each item on the assessment is scored by Arizona teachers or by Measurement, Inc. employees, according to scoring rubrics. The rubrics, which are printed in students' test booklets, are typically on a 1-3 or 1-4 scale (a score of **0** can be assigned as well) and are tailored to each assessment task. For example, one sample 3-point rubric for a mathematics assessment task is expressed as follows:

- *A **3** response shows an appropriate graph. It has a title, the axes labeled properly, and a scale.*

- *A **2** response shows a graph with minor problems resulting from errant calculations. For example, one endpoint of the graph may be incorrectly plotted. The axes are labeled properly and a scale is shown. Alternately, the response is correct in every way except that the axes are reversed (so that the line has a negative slope).*

- *A **1** response shows a graph with problems resulting from a flawed understanding of the problem. Both endpoints may be plotted incorrectly. The axes may not be labeled properly.*

- *Assign a **0** if the response shows no mathematical understanding of the problem (e.g., the slope of the line is negative when it should be positive), or if the student fails to respond to the question.*

A sample rubric appears in Appendix C.

Resource and Training Support

The Arizona legislature allocated no new funds to support the development and implementation of the Arizona Student Assessment Program. All work to develop and administer the assessments was supported under existing ADE testing budgets. According to the ADE Deputy Associate Superintendent, development and printing of the ASAP costs $1 million annually.

The Arizona Department of Education developed several activities and other training and communication mechanisms to educate teachers about the new assessment program and to keep them abreast of changes in it:

- In the Fall of 1990 Arizona conducted a statewide conference for teachers to introduce them to ASAP. Representatives from all of the state's 220 districts attended the conference.
- Over the course of the following school year, ADE officials traveled to various regional sites within the state to work further with teachers. The ADE conducted a "trainer of trainers" seminar during which ASAP's scoring rubrics and their application were taught to teachers, who were charged with returning to their districts to train their colleagues.
- Finally, the ADE has produced a videotape about ASAP and disseminated it to all districts; ADE also publishes a monthly/quarterly newsletter, *Measuring Up*, that updates teachers and administrators about developments in ASAP.

Interaction with Other Education Reforms

In addition to ASAP, other significant education reforms taking place in Arizona include the adoption of the state's *Essential Skills* and the requirement that every school district submit to the state a *District Assessment Plan* and an *Essential Skills Completion Report*.

Essential Skills

Arizona's curriculum frameworks, called the *Essential Skills*, were adopted during the 1986-87 school year and confirmed later by the Joint Legislative Committee on Goals for Educational Excellence. The state has introduced Essential Skills in the areas of language arts, mathematics, science, social studies, comprehensive health, foreign language, literature, music, dance, dramatic arts, and visual arts. As has been referenced above, the *Essential Skills* for twelfth-graders in reading and writing (two of the five components of the language arts *Essential Skills* — the others include speaking, listening, and language concepts) and for mathematics appear in Appendix B.

The ASAP is intended as an audit of students' progress toward attaining the *Essential Skills*; thus the two reforms are directly linked. Although ASAP currently assesses students' progress toward the *Essential Skills* in only language arts and mathematics, the ADE intends to expand ASAP to cover social studies and science in the future (a pilot is anticipated for 1996).

District Assessment Plan and Essential Skills Completion Report

Each school district must submit to the Arizona Department of Education a *District Assessment Plan*. This document sets forth "a plan of when and how districts will assess all *Essential Skills* in reading, writing and mathematics for grades K-3, 4-8, and 9-12. Districts may use ASAP performance-based assessment Forms A, B and/or C; their own CRTs; and/or portfolios. District CRTs and portfolios must match the state *Essential Skills* in content and process."[3]

The *District Assessment Plan* provides a framework for the district's annual *Essential Skills Completion Report*. In the *Essential Skills Completion Report*, the district reports the percentage of students mastering each *Essential Skill* in reading, writing, and math by the end of grades three, eight, and 12. Districts' first *Essential Skills Completion Reports* were due to the ADE in Summer 1994.

District Context

The Desert View Union High School District comprises nine high schools in northwest Phoenix and surrounding suburbs. The district consistently places among the top five districts in the state on standardized, multiple-choice exams (the ITBS and TAP), and its students achieved the highest average score on the twelfth grade ASAP in both 1993 and 1994.

The Desert View district is supportive of the ASAP and its goals. According to the district's Associate Superintendent:

> The system is evolving and certainly has considerable room for improvement. However, philosophically, the [district] has been and continues to be a strong believer in this type of systems approach to accountability. If true educational reform is to occur then high standards of student achievement must be clearly defined and linked to performance based assessment systems that hold schools accountable to their community for student achievement.

The Desert View Union Assessment Program

Desert View Union has an extensive, complex assessment program of its own, intended both to monitor student achievement and to hold teachers accountable for that achievement. The district administers criterion-referenced tests — both pre-tests and post-tests — in all subject areas and at all grade levels. In addition, the district administers its own performance assessments in most subject areas — social studies (U.S. history, world history, government), foreign languages, English (the "multiparagraph essay" described below), science, and in dance, the visual arts, and the theater arts. The district intends to expand its use of performance assessments to include all subjects (including mathematics) over the next couple of years.

---

[3] Bishop, C.D. (1993, August). *Assessment: Once Again Serving Teaching and Learning: The Arizona Student Assessment Program*. Phoenix: Arizona Department of Education, pp. 6-7.

The centerpiece of Desert View Union's performance assessment program is the "multiparagraph essay" assessment required of all students each year. Administration of this assessment takes approximately three days to administer to freshmen and sophomores and four days for juniors and seniors. Each year the type of assignment is different: freshmen write an "expository" essay, sophomores a "process" essay, juniors a "compare/contrast" essay, and seniors a "research" essay (seniors in Advanced Placement English are exempt). These essays are scored over the summer by the district's teachers. Desert View Union established this assessment in the late 1970s to monitor student writing skills and to guide writing instruction.

Accountability

Teachers interviewed for this case study stress that accountability is one important purpose of the district's assessment program. (The district implements a schoolwide merit pay system through which teachers obtain a bonus if certain goals are met each year; goals often include increases in test scores.) Although teachers acknowledge that accountability can have a positive impact on teaching and learning, they believe that the district looks at the test scores of teachers' students but fails to take into consideration the levels of ability of the students, considering actual, rather than gain, scores.

Teachers believe that students' ASAP scores ultimately will be used by Desert View Union for accountability purposes as well (though they are not currently). Given the likelihood that they will in the future be held accountable for their students' performance on ASAP, teachers stress the importance of designing the assessment tasks carefully; teachers suggest that, because they will teach to a test for which they are held accountable for student scores, the ADE should be certain that the skills being measured are the ones valued.

Resource and Training Support

The Desert View Union High School District has conducted some of its own training sessions with respect to ASAP. The district has developed materials, including practice assessment tasks, for use in classrooms, and it has conducted in-service workshops on the administration and scoring of ASAP for all teachers of twelfth-grade English and math. These district inservice sessions, occur each year just prior to the administration of ASAP. During the workshops, teachers receive and review practice assessments, share and discuss materials, and discuss the *Essential Skills*. Additionally, the district has remained involved in all stages of the introduction of ASAP, including serving as a scoring site, in part because by doing so teachers gain familiarity with the new assessment program.

Teachers at Manzanita High School were generally ambivalent about the training they received with respect to ASAP. Their comments included, "I'm sure I went through the training, but I don't remember anything about it," and "Yes. I think I participated." However, teachers who remembered the training said that they found the information directly pertinent to the mechanics (e.g., administration and scoring) of ASAP to be useful. They stressed, though, that they were already very comfortable with the idea of performance assessment because of the district's extensive performance assessment program.

School Context

This section of the case study principally focuses upon the impact ASAP has had on the Manzanita High School community. Although teachers at Manzanita High School are fans of performance-based assessment, their overall response to ASAP has been negative. Specifically, teachers question the quality of the assessment tasks, the validity of those tasks, and the absence of performance standards.

Impact on Teachers

Most teachers at Manzanita support both the concept of performance assessment and the forms through which it is implemented by the Desert View district. However, they are critical of the current ASAP assessment instruments. Criticisms fall into at least five categories:

- **The assessments are not valid**. Teachers do not believe that the assessment instruments assess student achievement across the breadth of the curriculum teachers are supposed to teach. Even allowing for the fact that the ASAP aims only to "audit" students' progress toward a subset of the state's *Essential Skills*, teachers dispute the validity of the instruments. One mathematics teacher said, "The math test is a reading comprehension test. It was not a traditional math test. It was technically poorly done, and the method for grading was poorly done. The test is invalid. The public is going to see results and think our kids didn't learn any math, and that's not true." He also said, "The twelfth grade math assessment covers only basic math skills (adding, subtracting, multiplying, and dividing). In 1994 one problem required that students graph a line, but no other 1994 and no 1993 problems required students to perform any algebra, geometry, trigonometry, or calculus."

- **No standards for performance have been established**. Though ASAP is aligned to the *Essential Skills*, the ADE has not yet established standards for performance on ASAP that would demonstrate student mastery of the *Essential Skills* assessed. One teacher in particular expressed concern about the fact that no standards have been set by either the ADE or Riverside: "If we have not decided beforehand what the students should know [referring to the level of performance that demonstrates competency], then what is the purpose of the test?"

- **The assessments, to date, have not been administered evenly**. Although ASAP is designed to allot students two hours for each section of the exam, teachers report that in 1993 and 1994 it was not administered evenly across schools and districts. As one teacher put it, "ASAP time constraints [during administration] are bizarre. We get faxes hourly [some of which modify time allowances] as we administer ASAP because so many questions come in to the ADE. No kid could do what was asked in the allotted time. . . . I don't believe in tests with uneven administration."

- **Most of the assessment booklets are scored out of state**. Teachers also expressed dissatisfaction with the fact that most of the assessments are scored out of state. Several teachers mentioned this flow of state funds out of the state as one of their major bones of contention with the ADE's administration of the assessment program.

- **The rubrics are too general**. Several teachers who have been involved with the ASAP scoring process reported difficulty using the rubrics. According to administrators in the Desert View district office, one of the district's recommendations to ADE after the pilot of ASAP was to make the rubrics more specific. They believe that the rubrics, which are specific to each task, are moving in the right direction but are not yet clear enough.

Teacher Reactions to Serving as ASAP Scorers

As has been mentioned, Desert View Union serves as a scoring site for ASAP. District and school administrators believe that the experience of scoring the assessments is a valuable professional development activity for teachers and that, by serving as a scoring site, the district is able to keep abreast of the evolution of ASAP: it is better to see what is coming from the inside than to have a *fait accompli* handed down from the state. However, teachers' responses to the scoring experience were decidedly negative. One teacher said, "I was involved in scoring once, and I never want to do it again." Another said, "I'm not putting myself through that again." When asked *why* they were uninterested in scoring again, they cited the rigidity and unresponsiveness of the Riverside trainers leading the scoring workshops and their own discomfort with the subjectivity of the scoring process.

The Combined Impact of ASAP and the District's Assessment Program

In addition to teachers' complaints particular to ASAP, they also note that the combined effect of ASAP and the district's extensive assessment program is to take a lot of instructional time out of the school year. The addition of ASAP to an already crowded and complex network of assessments means that Manzanita teachers must devote classroom time to prepare students for one more assessment and to administer the exam. In 1994, Desert View Union prepared a full-length practice assessment that all students took to prepare for the ASAP (in lieu of using ASAP Form A for practice). English and math teachers reported that they spent about one week of class time preparing their students for the ASAP, and administration of the exam took up most of another week (two hours on each of four days). Thus total time devoted to ASAP by English and math teachers of 12th graders was close to two weeks. All together, Manzanita teachers estimated that each student at the school spends about one month of the school year being tested.

English teachers lament that the time they now spend preparing their students for exams and administering the exams is time they once spent teaching literature. One teacher reported that he had to cut *The Grapes of Wrath*, *The Adventures of Huckleberry Finn*, *Of Mice and Men*, and *The Great Gatsby* from his curriculum because state and district assessments require him to spend so much time on other activities. Other teachers commiserate, one stating that she now tends to teach short stories instead of novels and another saying that she still tries to "sneak" some poetry into her 12th grade English curriculum. What is more, teachers wonder to what end the state and district assessments have encouraged them to modify instruction. Teachers say that they have modified curriculum and instruction to accommodate the assessments, but they do not believe that their students are learning any better than they did formerly. In the words of one teacher, "The emphasis in this district on assessment has made me a different teacher, not necessarily a better one."

This impact on teachers, it should be stressed again, is a function not of ASAP alone but of ASAP and the district's assessment program. The district administers 14 performance assessments to ninth, tenth, and eleventh graders each year. As one teacher put it, "You can't teach in this district if you don't believe in assessment."

Impact on Students

Students, too, were critical of their experience with ASAP. Two honors students remarked that the essay they were asked to read for the ASAP reading assessment was poorly written, illogical, and self-contradictory. Additionally, one of these students said of the reading and writing assessments, "Ironically, the AP English students had the hardest time with the time limit because they had the most to say." The other commented, "Higher level writing is not what ASAP required. I had to retrain myself to do what they wanted." These two students remarked that they did not find the ASAP to be challenging, nor did their classmates in the honors program; they suggested that ASAP was not a good assessment of their abilities or performance levels but that it may have been appropriate for students at lower performance levels.

One young man interviewed said that, while he did not enjoy the ASAP, he could see the relevancy to his future of the skills it tested. He intended to join the U.S. Navy's Search and Rescue division upon high school graduation, and said of ASAP's writing section, "Writing is hard for me, but it's relevant to my future. I'll have to write a lot of reports and things like that."

Another student, a boy who struggles with his school work, failed to see any relevance of ASAP to his life. "It's stupid. We don't need to do that stuff." When asked about the purpose of ASAP, he said, "Do we have knowledge about cars or something [the mathematics assessment tasks — on the theme of consumer decision-making — focused on choosing to purchase a car based on various criteria]? I really don't know."

The one special education student interviewed for this case study did not participate in ASAP, as his IEP exempted him from participation. However, this student, despite his learning disability, was college-bound.

Impact on Parents

The two parents interviewed were not equally well informed about ASAP. One parent knew that her daughter had taken the exam, but she knew little about the examination, itself. She expressed an interest in receiving more information from the school about ASAP.

As a member of Manzanita's site-based planning team, the other parent was better informed about ASAP. In her opinion, ASAP is a "minimum competency exam" which tells her nothing about her children's performance that she doesn't learn from their grades and their performance on AP exams. (She noted that her son, who graduated from Manzanita in 1993, was "amused" to score a perfect "5" on the AP calculus exam but to miss points on the ASAP math exam.)

This parent hypothesized a connection between the district's top performance on ASAP and the district's own assessment program: "The district's writing assessment has changed the way teachers teach, helping students' performance on ASAP." She also believes her children have been taught more about how to write because of the district's assessment program.

Future Plans

The Arizona Department of Education intends to add assessments to ASAP in science and social studies over the next couple of years; a pilot is currently anticipated for 1996. In addition, satisfactory performance on the 12th grade ASAP will become a requirement for graduation. Originally, this plan was to have gone into effect in 1996, but the ADE now plans to institute ASAP as a graduation requirement for the class of 2004.

Although this expansion of ASAP is still part of the ADE's plans as they currently stand, in January 1995 the Department's new superintendent temporarily postponed ASAP so that the program as it now exists could be revisited. Thus, ASAP was not administered in winter 1995. The Superintendent intends to reinstate ASAP after the program has been reviewed and, in all likelihood, modified.

Conclusions

The direct impact exerted by ASAP on teachers and students at Manzanita High School was marginal. Though their responses to the state-mandated assessment ranged from the ambivalent to the decidedly negative, overall teachers suggested that ASAP has not yet had a significant effect on teaching and learning in their classrooms.

Teachers' comments about ASAP clearly demonstrated their dissatisfaction with several aspects of the development and implementation of the assessment program. According to their comments:

- ASAP was developed by experts under contract to the Arizona Department of Education, and teacher involvement was minimal (as perceived by teachers);
- Standards for individual student performance have not yet been articulated; and
- The assessments were not valid because they did not assess students on the full range of desired skills within the assessed subject areas.

Although ASAP itself is not popular among Manzanita teachers and students, performance assessment, both in theory and in other manifestations, is not merely popular but a well-entrenched feature of the Desert View Union educational program. Indeed, teachers in the highest-performing district on the ASAP are, arguably, uniquely positioned to critique the program: despite their district's impressive showing, these teachers are sharply critical of the flaws they perceive in the ASAP system. If ASAP is to deliver to Arizona that which its creators envisioned — a system for measuring and certifying students achievement of the *Essential Skills* — it would seem that the state might benefit from revisiting the assessment in view of teacher concern.

APPENDIX A

Mathematics Task

In your industrial arts class, you are making projects with wood. You and three of your friends — Jan, Keisha, and Earl — decide to make birdhouses for your first project.

Exercise A

Your teacher, Mr. Ramírez, gives the four of you one piece of wood 64 inches long and gives each of you a slip of paper that reads. "Measure the board and cut off ¼ of it."

- First, Jan measures the board and cuts off ¼ of it. She gives the rest of the board to Keisha.
- Next, Keisha measures the board and cuts off ¼ of it. she gives the rest of the board to Earl.
- Finally, Earl measures the board and cuts off ¼ of it. He gives the rest of it to you.

You discover that you have much more wood than you need. Keisha and Earl complain that they don't have enough.

Explain what was confusing about Mr. Ramirez's directions and tell how he could have made them clearer. Use the Review Check List below to help make your answer complete.

©1992 The Riverside Publishing Company. All rights reserved

Review Check List

- ☐ This explanation tells what was confusing about Mr. Ramírez's directions.
- ☐ This explanation tells how Mr. Ramírez could have made his directions clearer.
- ☐ The ideas in this explanation follow a logical order.
- ☐ This explanation contains complete sentences and correct spelling, capitalization, and punctuation. It is neatly written.

©1992 The Riverside Publishing Company. All rights reserved.

APPENDIX B

Essential Skills

Essential Skills: Reading, Grades 9-12

1. reads silently daily
2. recalls and states prior knowledge about material
3. predicts content
4. states purpose or goal for reading
5. plans strategy for reading
6. reads fluently
7. summarizes, clarifies, predicts, and asks questions while reading
8. monitors own comprehension and self-corrects when necessary
9. reads and demonstrates comprehension of a personal experience narrative
10. reads and demonstrates comprehension of fiction
11. reads and demonstrates comprehension of an informative report
12. reads and demonstrates comprehension of a form of communication
13. reads and demonstrates comprehension of a poem
14. reads and demonstrates comprehension of a summary
15. reads and demonstrates comprehension of an essay
16. reads and demonstrates comprehension of persuasion
17. reads and demonstrates comprehension of a review/evaluation/critique

Essential Skills: Writing, Grades 9-12

1. participates in some aspect of the writing process daily
2. participates in a prewriting activity
3. identifies purpose for writing
4. identifies audience
5. narrows topic/subject
6. decides on method of development
7. free writes to get first draft
8. monitors own writing
9. writes next draft
10. revises for ideas
11. revises for paragraph structure
12. revises for sentence structure
13. revises for word choice
14. edits for usage
15. edits for capitalization, punctuation and spelling
16. rewrites paper
17. proofreads final draft
18. shares or publishes final draft
19. writes a personal experience narrative
20. writes a short story or play
21. writes a report
22. writes a communication
23. writes a poem
24. writes a summary/precis
25. writes a specialized expository paper (essay)

26. writes a persuasive paper
27. writes an evaluation/critique

Essential Skills: Mathematics, Grades 9-12

1. use the four arithmetic operations on rational numbers (expressed as common fractions, decimal fractions or percents) with accuracy and reasonable speed, using pencil and paper or calculator
2. perform mental calculations of the four operations with appropriate levels of precision
3. understand the concept of ordering the real numbers by locating their corresponding points on the number line
4. write mathematical expressions and sentences to replace verbal expressions, using appropriate operation, number and relation symbols
5. use standard rules for order of operations, together with symbols of inclusion to correctly evaluate numerical expressions and formulas in a variety of problem settings
6. understand and use properties of equality and inequality
7. understand the approximate nature of measurement and apply the precision of measurements to the resulting accuracy of calculations
8. make unit conversions with appropriate levels of accuracy between and within standard systems of measures
9. use appropriate standard units of measurement to make reasonable estimates of linear, area, volume and weight measures of objects commonly encountered in daily life
10. select and use appropriate formulas or procedures to determine indirectly a measure when a direct measurement is not available or feasible
11. identify and distinguish between geometric elements, such as lines, angles, planes, polygons, circles and regular solids
12. understand and apply basic geometric relationships such as parallel, intersecting, perpendicular, similarity, proportionality and congruence
13. calculate areas of regular polygons and circles and volumes of rectangular solids, cylinders and spheres
14. use compass, protractor and ruler to perform standard geometric constructions
15. apply the Pythagorean theorem in solving problems
16. select and use appropriate principles of counting collections and arrangements of objects or outcomes of sequential procedures
17. select and use appropriate statistical measures to describe sets of data
18. explain variability of statistical measures in terms of sampling process or population differences
19. use the concept of mathematical expectation to estimate outcomes of random processes
20. identify and explain misuses of statistics
21. use experimental observations to estimate empirical probabilities and population parameters
22. distinguish between independent and dependent events and use conditional probabilities
23. add, subtract, multiply and divide with monomials and binomials
24. simplify and evaluate algebraic expressions which involve integral exponents and square roots
25. simplify rational algebraic expressions having monomial denominators
26. solve linear equations and inequalities in one variable, and apply these to solve problems

27. solve systems of linear equations and inequalities in two variables, graphically and analytically, and formulate such systems to solve problems
28. solve problems using ratio and proportion, percent or direct and inverse variation
29. identify and distinguish between arithmetic and geometric progressions and state a rule or formula for the general term of such a progression
30. write the linear relationship between two variables by inspection of its graph or of the set of its ordered pairs
31. graph the inverse of a function by inspection of the graph of the function
32. identify and explain graphic misrepresentation or distortions of sets of data
33. identify, interpret and construct graphs of nonlinear relations such as parabolas and circles
34. distinguish between inductive and deductive reasoning and explain how each is most appropriately used
35. recognize when conditions of definitions are satisfied
36. use deductive reasoning to generate conclusions
37. use statistical reasoning to test hypotheses informally
38. identify valid and invalid arguments

APPENDIX C

Mathematics Rubrics

Rubric for Exercise A, Observation 2

Explain what was confusing about Mr. Ramírez's directions and tell how he could have made them clearer.

Note: Acceptable responses should address the following two parts of the problem:

- What was confusing about the directions (i.e., The students did not realize that Mr. Ramírez meant ¼ of the whole board.)

- How Mr. Ramírez could have made the directions clearer (i.e., He could have suggested that the students measure the board and divide it into fourths before doing any cutting.)

A **3** response points out that *each* student should not have measured the board given to him or her and determined what ¼ of it would be. The board should have been measured once, before any cutting was done, and divided into fourths. Mr. Ramírez could have asked one person to measure the board and divide it into fourths and the other three to make the cuts. Other reasonable methods of dividing the work among the student should be accepted.

A **2** response provides a clear explanation of either what was confusing about Mr. Ramírez's directions or what could be done to make them clearer, but provides a vague explanation of the other part.

A **1** response explains only one part of the problem (i.e., either what was confusing about Mr. Ramírez's directions or what could be done to make them clearer). Alternately, the student provides a vague or incomplete explanation of both parts of the problem.

Assign a **0** if the response shows no mathematical understanding of the problem or if the student fails to respond to the question.

Assign N/S (Not Scorable) if the response is illegible or unreadable.